

STIC-Biotech/ChemLib

169433

ME

From: Chan, Christina
Sent: Monday, October 24, 2005 5:39 PM
To: Walicka, Malgorzata; STIC-Biotech/ChemLib
Subject: RE:

Please rush. Thanks

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Walicka, Malgorzata
Sent: Monday, October 24, 2005 3:50 PM
To: Chan, Christina
Subject:

Christina, please authorize rush search of 3 short poeptides of SEQ ID NO:2, 3, and 9 in applicaiton 09/774,954.
Please also against DNA data bases. Thank you.

Malgorzata

Malgorzata A. Walicka, Ph.D.
Patent Examiner
Art Unit 1652, Recombinant Enzymes
USPTO, Remsen Building, Room 2C76
400 Dulany St.
Alexandria, VA 22313
Mail Room 2C70
Tel. (571) 272-0944, fax (571) 273-0944

2aa 365
3 / 61
9aa 61

ME

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 210.606 seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLPCWGR.....QGRPSFFCMRPPKLRDEF 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	100.0	388	1	Q9H488 homo sapien
2	1979	100.0	388	1	Q6EV69 pan troglod
3	1895	95.8	391	2	Q7YR27 bos taurus
4	1878	94.9	391	2	Q6EV66 bos taurus
5	1854	93.7	395	2	Q6EV70 rattus norv
6	1853	93.6	378	2	Q7YR23 sus scrofa
7	1832	92.6	393	1	Q7YR23 sus scrofa
8	1829	92.4	393	2	Q8C8E4 mus musculus
9	1672	84.5	380	2	Q8A8B4 gallus gall
10	1586	80.1	351	2	Q7YR66 bos taurus
11	1490	75.3	380	2	Q6A0S0 xenopus tro
12	1490	75.3	396	2	Q6EV71 xenopus tro
13	1474	74.5	395	2	Q7T028 brachydanio
14	1447	73.1	380	2	Q8AXS8 xenopus lae
15	1427	72.1	384	2	Q70AG7 fugu rubrip
16	1395	70.5	387	2	Q6EV65 tetracodon n
17	848.5	42.9	381	2	Q6EV67 ciona savig
18	838	42.3	402	2	Q6S9S1 drosophila
19	831	42.0	402	1	QFUI DROME
20	827	41.8	402	2	Q86S77 drosophila
21	789.5	39.9	399	2	Q7QHS7 anopheles g
22	786	39.7	353	2	Q6S9S0 bombyx mori
23	778	39.3	364	2	Q8MSR1 drosophila
24	761	38.5	402	2	Q6EV68 drosophila
25	660.5	33.4	381	1	QFUI CAEEL
26	644	32.5	434	2	Q6S9S2 caenorhabdi
27	578.5	29.2	385	2	Q6S9R9 onchocerca
28	295	14.9	61	1	QFUI CRIGR
29	182.5	9.2	424	1	QFUI CAEEL
30	161.5	8.2	490	2	Q6S9L6 drosophila
31	159.5	8.1	423	2	Q7K729 drosophila

RESULT 1

ID	OFUI_HUMAN	STANDARD;	PRT;	388 AA.
AC	Q9H488; Q14685; Q9BW76;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)			
DE	(Peptide O-fucosyltransferase) (O-FUCT-1).			
GN	Name=POFUT1; Synonyms=KIAA0180;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.			
RC	TISSUE=Heart;			
RC	MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;			
RA	Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,			
RA	Haltiwanger R.S.,			
RT	"Modification of epidermal growth factor-like repeats with O-fucose:			
RT	molecular cloning and expression of a novel GDP-fucose protein O-			
RT	fucosyltransferase."			
RL	J. Biol. Chem. 276:40338-40345(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Brain;			
RC	MEDLINE=22158633; PubMed=12168954;			
RA	Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.,			
RT	"Construction of expression-ready cDNA clones for KIAA genes: manual			
RT	curation of 330 KIAA cDNA clones."			
RL	DNA Res. 9:99-106(2002).			
RN	[3]			
RP	SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).			
RC	TISSUE=Bone marrow;			
RC	MEDLINE=96281124; PubMed=8724849;			
RA	Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. V.			
RT	The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by			
RL	analysis of cDNA clones from human cell line KG-1."			
RL	DNA Res. 3:17-24(1996).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;			
RA	Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Conn R.E.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			

Q9W589 drosophila
Q6EV55 caenorhabdi
Q6EV53 fugu rubrip
Q6EV54 drosophila
Q6EV56 pan troglod
Q7Z4N0 homo sapien
Q8VHI3 mus musculu
Q7YZS7 ciona intes
Q7LNG6 gallus gall
Q6EV52 ciona savig
Q7YR55 bos taurus
Q7PRQ3 anopheles g
Q6S9L5 bombyx mori
Q9MA87 arabidopsis

ALIGNMENTS

THIS PAGE BLANK (USPTO)

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.P.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levaehalaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McMay K., Mullikin J.C., Murray A.A.,
 RA Mine S.A., Mistry D., Moore M.J.F., Mullikin J.C., McMurray A.A.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 RA Suze C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,
 RA Swann R.D., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.,
 RT "The DNA sequence and comparative analysis of human chromosome 20."
 RL Nature 414:865-871(2001).
 RP (5)
 RPN SEQUENCE FROM N.A. (ISOFORM 2).
 RRC TISSUE=Brain;
 RRX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L.H., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters K.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences".
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [6]
 RPN CHARACTERIZATION.
 MEDLINE=97175972; PubMed=9023546;
 RA Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
 RA "Identification of a GpL-fucose:polypeptide fucosyltransferase and
 enzymatic addition of O-linked fucose to EGF domains".
 Glycobiology 6:837-842(1996).
 RPN -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
 O-glycosidic linkage to a conserved serine or threonine residue in
 EGF domains.
 RPN -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
 beta-L-fucose to the serine hydroxy group of a protein acceptor.
 RPN -!- COFACTOR: Manganese.
 RPN -!- PATHWAY: Glycosylation.
 RPN -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1;
 IsoId=Q9H488-1; Sequence=Displayed;
 Name=2;
 IsoId=Q9H488-2; Sequence=YSP_001809;
 RPN Note=No experimental confirmation available;
 RPN -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
 lung, liver, skeletal muscle, kidney and pancreas.
 RPN -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.

 RPN This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

THIS PAGE BLANK (USPTO)

|||||
Db 1071 CGGAGCGGACCTCCAGGAGCGCCGCTCTTCTTCTCGGACGAGCCGCCCTTAAG 1130
|||
Qy 361 LeuArgaspGluPhe 365
|||
Db 1131 CTGCGGACGAGTTC 1145
|||
RESULT 10
Db0002
LOCUS D80002 5189 bp mRNA linear PRI 10-JAN-2003
DEFINITION Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION D80002 GI:20521837
VERSION D80002.2
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Nagase, T., Seki, N., Ishikawa, K., Tanaka, A. and Nomura, N.
Prediction of the coding sequences of unidentified human genes. V.
The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
analysis of cDNA clones from human cell line KG-1
DNA Res. 3 (1), 17-24 (1996)
96281124
8724849
2
Chiang, P.W., Wang, S., Smithivas, P., Song, W.J., Ramamoorthy, S.,
Hillman, J., Puett, S., Van Keuren, M.L., Cronbez, E., Kumar, A.,
Glover, T.W., Miller, D.B., Tsai, C.H., Blackburn, C.C., Chen, X.N.,
Sun, Z., Cheng, J.F., Korenberg, J.R. and Kurnit, D.M.
Identification and analysis of the human and murine putative
chromatin structure regulator SUPT6H and Supt6h
Genomics 34 (3), 328-333 (1996)
96374824
8786132
3 (bases 1 to 5189)
Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.
Direct Submission
Submitted (12-DEC-1995) Oasamu Ohara, Kazusa DNA Research Institute;
152-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)
On May 9, 2002 this sequence version replaced gi:1136419.
Location/Qualifiers
1. 5189
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ha02567s1"
/sex="male"
/cell_line="KG-1"
/cell_type="myeloblast"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/note="This sequence was obtained by subcloning of the DNA
fragments derived from two cDNA clones (1 - 1066 was
derived from pf00933 and 1067 - 5189 was derived from
ha02567)."
1. 5189
/gene="KIAA0180"
1. 1213
/gene="KIAA0180"
/citation=[2]
/codon_start=2
/protein_id="BA011497.2"
/db_xref="GI:20521838"
/translation="PPRLCAAAGSGSRADMGAAWAPLSVSLILLPLPFGPAGSW
DPAGVLLYCPGMRFGNADHFLGSLAFKLLNLTAVPPMIEYOHKPPPTNLHVS
QXYKLEPLOAYHRVISELDEMEKLAHPPEKEVAVCFEVAQSRDPDKTCPMKEG
NPFGEFDQFVFNKSELSFTGISFASYSYRQWSQRFKHPVLALPGAPQPVLE
EHRFLQKYMWSDEKVTGEAQIHAHLVRYPIYVGHIRIGSDWKNACAMLDKGTAGSHF
MASPCQVYSRSTAPLWTMCLPDLKEIQRAVKLWRSLSAQSYVATDESYPVEL
QQLFKGKVKVSLKPEVAQVDLYILGQADHDFIGNCVSFTTAFVKRERDLQGRPSFFG

MDRPPKLRDEF"
Alignment Scores:
Pred. No.: 6.7e-174 Length: 5189
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-774-954-2 (1-365) x D80002 (1-5189)
Qy 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||
Db 116 ATGCTCGGGGCTCTCGGAGCCGCGGTTACTGCTCTACTCCCTCGATGGGGCGC 175
|||
Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||
Db 176 TTTGGGAACAGGCGCGATCACTTCTTGGGCTCTCTGGCAATTGGCAAGCTCTAAACCGT 235
|||
Qy 41 ThrLeuAlaValProTrpPileGluTyrGlnHisHisLysProProPheThrAsnLeu 60
|||
Db 236 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCACAGGCTCTCTTTCAACAACCTC 295
|||
Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
|||
Db 296 CATGTGCTCTACCAGAAAGTATTCAAGCTGGAGCCCTCCAGGCTTACCATCGGCTCATC 355
|||
Qy 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
|||
Db 356 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCACTGCGCCCTCGAGAGCGGGTG 415
|||
Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
|||
Db 416 GCATCTGCTTTGAGTGGCAGCCAGGAGCCAGATAGAGACGTGCCCATGAAG 475
|||
Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
|||
Db 476 GAAGGAACCCCTTTGGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 535
|||
Qy 141 LeuPheThrGlyLysSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
|||
Db 536 CTTTTCACAGGCAATTTCTTCAGTGCTTCTTACAGAGAACATGGAGCCAGAGATTTCT 595
|||
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
|||
Db 596 CCAAGGAGACATCCGGTGTCTTGGCTGCCAGAGCCCGCCAGCTTCCCGTCTCTAGAG 655
|||
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
|||
Db 656 GAACACAGGCCCACTACAGAGTACATGTTATGGTCAGACGAAATGGTGAAGACGGGAGAG 715
|||
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
|||
Db 716 GCCCAGATTTACATGCCACCTTTGTCGGCCCTATGTGGGCATTCATCTGCGCATTTGCTCT 775
|||
Qy 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
|||
Db 776 GACTGGAAGAACCGCTGTGCCATCTGTGAGAGCGGAGCTGCAGGCTCGCACCTTCATGGCC 835
|||
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
|||
Db 836 TCTCCGCAAGTGTGTGGGCTACAGCGGACACAGCGGCGCCCTCCATCAGTACTATGTGC 895
|||
Qy 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
|||
Db 896 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 955
|||
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
|||
Db 956 CAGTCGGTGTACGTTGCTACTGATTTCCGAGAGTTATGTGCTCAGCTCCACAGCTCTTC 1015
|||

THIS PAGE BLANK (11SPT0)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 210.606 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLCPMGR.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	100.0	388	1	Q9H488 homo sapien
2	1979	100.0	388	1	O6ev69 pan troglod
3	1895	95.8	391	2	Q7YrE7 bos taurus
4	1878	94.9	391	2	Q6EV66 bos taurus
5	1854	93.7	395	2	Q6EV70 rattus norv
6	1853	93.6	378	2	Q7YrZ3 sus scrofa
7	1832	92.6	393	1	QFUI MOUSE
8	1829	92.4	393	2	Q8C8F4 mus musculu
9	1672	84.5	380	2	Q8AWB4 gallus gall
10	1586	80.1	351	2	Q7YrE6 bos taurus
11	1490	75.3	380	2	Q640S0 xenopus tro
12	1490	75.3	396	2	Q6EV71 xenopus tro
13	1474	74.5	395	2	Q7T028 brachydanio
14	1447	73.1	380	2	Q8AXS8 xenopus lae
15	1427	72.1	384	2	Q70AG7 fugu rubrip
16	1395	70.5	387	2	Q6EV65 tetraodon n
17	848.5	42.9	381	2	Q6EV67 ciona savig
18	838	42.3	402	2	Q659S1 drosophila
19	831	42.0	402	1	QFUI DROME
20	827	41.8	402	2	Q86SA7 drosophila
21	789.5	39.9	399	2	Q7QHS7 anopheles g
22	786	39.7	353	2	Q659S0 bombyx mori
23	778	39.3	364	2	Q8MSR1 drosophila
24	761	38.5	402	2	Q6EV68 drosophila
25	660.5	33.4	381	1	QFUI CABEL
26	644	32.5	434	2	Q859S2 caenorhabdi
27	578.5	29.2	385	2	Q659R9 onchocerca
28	295	14.9	61	1	QFUI CRIGR
29	182.5	9.2	424	1	Q8WR51 caenorhabdi
30	161.5	8.2	490	2	Q659I6 drosophila
31	159.5	8.1	423	2	Q7K729 drosophila

RESULT 1

ID	OFUL_HUMAN	STANDARD;	PRT;	388 AA.
AC	Q9H488; Q14685; Q9BW76;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)			
DE	(Peptide O-fucosyltransferase) (O-FUCT-1).			
GN	Name=POFUT1; Synonyms=KIAA0180;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
OX	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.			
RC	TISSUE=Heart;			
RX	MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;			
RX	Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,			
RA	Haltiwanger R.S.;			
RA	"Modification of epidermal growth factor-like repeats with O-fucose:			
RT	molecular cloning and expression of a novel GDP-fucose protein O-			
RT	fucosyltransferase.";			
RT	J. Biol. Chem. 276:40338-40345(2001).			
RL	[2]			
RN	SEQUENCE FROM N.A. (ISOFORM 1).			
RP	TISSUE=Brain;			
RX	MEDLINE=22158633; PubMed=12168954;			
RX	Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;			
RA	"Construction of expression-ready cDNA clones for KIAA genes: manual			
RT	curation of 330 KIAA cDNA clones.";			
RT	DNA Res. 9:99-106(2002).			
RL	[3]			
RN	SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).			
RP	TISSUE=Bone marrow;			
RX	MEDLINE=96281124; PubMed=8724849;			
RX	Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;			
RA	"Prediction of the coding sequences of unidentified human genes. V.			
RT	The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by			
RT	analysis of cDNA clones from human cell line KG-1.";			
RL	[4]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;			
RX	Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,			
RA	Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			

Q9W589 drosophila
Q6ev55 caenorhabdi
Q6ev53 fugu rubrip
Q6ev54 drosophila
Q6ev56 pan troglod
Q7z4n0 homo sapien
Q8vhi3 mus musculu
Q7yze7 ciona intes
Q7tin6 gallus gall
Q6ev52 ciona savig
Q7yrr5 bos taurus
Q7prq3 anopheles g
Q659I5 bombyx mori
Q9ma87 arabisopsis

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivaeslao M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore A.C., Vaudin M., Wall M., Wallis J.M.,
RA Tracey A., Tromans N.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RA "The DNA sequence and comparative analysis of human chromosome 20."
RT Nature 414:865-871 (2001).
RL [5]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RP TISSUE=Brain;
RC MEDLINE=2398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.T., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]
RP CHARACTERIZATION
RX MEDLINE=91719372; PubMed=9023546;
RA Wang Y., Lee G.F., Kelley R.F., Spellman M.W.
RT "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT enzymatic addition of O-linked fucose to EGF domains."
RL Glycobiology 6:837-842 (1996).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains.
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese.
CC -!- PATHWAY: Glycosylation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9H488-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9H488-2; Sequence=VSP_001809;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC lung, liver, skeletal muscle, kidney and pancreas.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AF375884; AAL09576.1; --
DR EMBL; D80002; BAA11497.2; --
DR EMBL; AL121897; CAC16424.1; --
DR EMBL; BC00582; AAH00582.1; --
DR Genew; HGNC:14988; POFUT1.
DR H-InvdB; HIX0015724; --
DR MIM; 607491; -- C:cytoplasm; ISS.
DR GO; GO:0005737; C:cytoplasm; ISS.
DR GO; GO:0030173; C:integral to Golgi membrane; NAS.
DR GO; GO:0046922; F:peptide-O-fucosyltransferase activity; TAS.
DR GO; GO:0009790; P:premyon development; NAS.
DR GO; GO:0007219; P:Notch signaling pathway; ISS.
DR GO; GO:0016266; P:O-glycan processing; TAS.
DR GO; GO:0016266; P:O-glycan processing; TAS.
DR GO; GO:0045449; P:regulation of transcription; NAS.
DR GO; GO:0046445; P:regulation of translation; NAS.
DR GO; GO:0046445; P:regulation of translation; NAS.
KW Alternative splicing; Fucose metabolism; Glycoprotein;
KW Glycosyltransferase; Manganese; Signal; Transferase.
FT SIGNAL 1 26 Potential.
FT CHAIN 27 388 GDP-fucose protein O-fucosyltransferase
FT FT 1.
FT CARBOHYD 62 62 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 182 388
FT FT
FT SPQCVGYSRSTAAPLTWTMCLPDLKEIOAKVLRSLDAGSVVATDSVYVPELQQLF 300
FT SVYVATDSVYVPELQQLFAGKVKVSVSLKPEVAOVDLYLG
FT QADHFIQGVSSFTAFVVRERDLQRPSPFGMDRPPKLRD
FT EF -> RENNHCVTLLPFR (in isoform 2).
FT FT
FT SEQUENCE 388 AA; 43955 MW; 3FACCCA34D02415 CRC64;
FT FT
FT Query Match 100.0%; Score 1979; DB 1; Length 388;
FT Best Local Similarity 100.0%; Pred. No. 1.1e-160;
FT Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MPAGSWDPAGVLLYCPMGRCGNQADHFLGSLAFKLLNRTTAVPPWIEYQHKPPFTNL 60
Db 24 MPAGSWDPAGVLLYCPMGRCGNQADHFLGSLAFKLLNRTTAVPPWIEYQHKPPFTNL 83
Qy 61 HVSQYKFKLEPLOAHVRVVISLEDFMEKLAETHPPKRVAYCFEVAOORSDDKKTCPMK 120
Db 84 HVSQYKFKLEPLOAHVRVVISLEDFMEKLAETHPPKRVAYCFEVAOORSDDKKTCPMK 143
Qy 121 EGNPPGPPWQDFHVSFNKSELFTGISFSASYREQWSQRFSPKSHPVLLALPGAPQFPVLE 180
Db 144 EGNPPGPPWQDFHVSFNKSELFTGISFSASYREQWSQRFSPKSHPVLLALPGAPQFPVLE 203
Qy 181 EHPLOKYMWSDEMYKTEGAOIHAAHLVRPVYVGIHLRIGSDWKNAKMLKDGTAGSHFWA 240
Db 204 EHPLOKYMWSDEMYKTEGAOIHAAHLVRPVYVGIHLRIGSDWKNAKMLKDGTAGSHFWA 263
Qy 241 SPQCVGYSRSTAAPLTWTMCLPDLKEIOAKVLRSLDAGSVVATDSVYVPELQQLF 300
Db 264 SPQCVGYSRSTAAPLTWTMCLPDLKEIOAKVLRSLDAGSVVATDSVYVPELQQLF 323
Qy 301 KGGKVVVSLKPEVAQVDLYLTGQADHFIQGVSSFTAFVVRERDLQRPSPFGMDRPPK 360
Db 324 KGGKVVVSLKPEVAQVDLYLTGQADHFIQGVSSFTAFVVRERDLQRPSPFGMDRPPK 383
Qy 361 LRDEF 365
Db 384 LRDEF 388
RESULT 2
ID OFUL PANTR STANDARD; PRT; 388 AA.
AC Q9EV69;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)

DE (Peptide O-fucosyltransferase) (O-FucT-1).

GN Name-POFUT1; Synonyms-FUT12;

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OC NCBI_TaxID=9598;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=12966037; DOI=10.1093/glycob/cwg113;

RA Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C., Oriol R.

RT "A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs."

RL Glycobiology 13:1C-5C(2003).

CC -1- FUNCTION: Catalyzes the reaction that attaches fucose through an O-glycosidic linkage to a conserved serine or threonine residue in EGF domains (By similarity).

CC -1- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-beta-L-fucose to the serine hydroxy group of a protein acceptor.

CC -1- COFACTOR: Manganese (By similarity).

CC -1- PATHWAY: Glycosylation.

CC -1- SIMILARITY: Belongs to the glycosyltransferase 68 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@sib.ac.uk).

CC EMBL; AJ781500; CA03712.1; -

DR Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;

KW Signal; Transferase.

FT SIGNAL 1 26 Potential.

FT CHAIN 27 388 GDP-fucose protein O-fucosyltransferase

FT CARBOHYD 62 62 1- N-linked (GlcNAc...) (Potential).

FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).

FT SEQUENCE 388 AA; 43955 MW; 3FACCCA434D02415 CRC64;

Query Match 100.0%; Score 1979; DB 1; Length 388;

Best Local Similarity 100.0%; Pred. No. 1.1e-160; Indels 0; Gaps 0;

Matches 365; Conservative 0; Mismatches 0;

QY 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPPTNL 60

DB 24 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPPTNL 83

QY 61 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAARSPDKKTCPMK 120

DB 84 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAARSPDKKTCPMK 143

QY 121 EGNPPGPFQDFHVSFNKSELFTGISFSASVREQMSORFSPKEHPVLALPCAPQFPVLE 180

DB 144 EGNPPGPFQDFHVSFNKSELFTGISFSASVREQMSORFSPKEHPVLALPCAPQFPVLE 203

QY 181 EHRPLQKYMWSDEMVTGEAQIHAHLVPRVPGIHLRIGSDWKNAKAMLDKGTAGSHFMA 240

DB 204 EHRPLQKYMWSDEMVTGEAQIHAHLVPRVPGIHLRIGSDWKNAKAMLDKGTAGSHFMA 263

QY 241 SPQCVGYSRSTAAPLTMTCLPDLKEIQRAVKLVWRSIDQSVVATDSSEYVPELQOLF 300

DB 264 SPQCVGYSRSTAAPLTMTCLPDLKEIQRAVKLVWRSIDQSVVATDSSEYVPELQOLF 323

QY 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360

DB 324 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 383

QY 361 LRDEF 365

DB 384 LRDEF 388

RESULT 3

Q7YRE7 PRELIMINARY; PRT; 391 AA.

ID Q7YRE7

AC Q7YRE7

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Protein O-fucosyltransferase 1a.

GN Name: pofut1;

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovinae; Bos.

OC NCBI_TaxID=9913;

OX [1]

RN SEQUENCE FROM N.A.

RP Lorient C., Germot A., Dupuy F., Maftah A.;

RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AX144580; AA02332.1; -

DR GO; GO:0016757; F:transferase activity, transferring glycosyl...; IEA.

KW Glycosyltransferase; Transferase.

SQ SEQUENCE 391 AA; 44307 MW; 96762AB81A2027AD CRC64;

Query Match 95.8%; Score 1895; DB 2; Length 391;

Best Local Similarity 93.4%; Pred. No. 1.8e-153; Indels 0; Gaps 0;

Matches 341; Conservative 17; Mismatches 7;

QY 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPPTNL 60

DB 27 LPVGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPPTNV 86

QY 61 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAARSPDKKTCPMK 120

DB 87 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAARSPDKKTCPMK 146

QY 121 EGNPPGPFQDFHVSFNKSELFTGISFSASVREQMSORFSPKEHPVLALPCAPQFPVLE 180

DB 147 EGNPPGPFQDFHVSFNKSELFTGISFSASVREQMSORFSPKEHPVLALPCAPQFPVLE 206

QY 181 EHRPLQKYMWSDEMVTGEAQIHAHLVPRVPGIHLRIGSDWKNAKAMLDKGTAGSHFMA 240

DB 207 EHRPLQKYMWSDEMVTGEAQIHAHLVPRVPGIHLRIGSDWKNAKAMLDKGTAGSHFMA 266

QY 241 SPQCVGYSRSTAAPLTMTCLPDLKEIQRAVKLVWRSIDQSVVATDSSEYVPELQOLF 300

DB 267 SPQCVGYSRSTAAPLTMTCLPDLKEIQRAVKLVWRSIDQSVVATDSSEYVPELQOLF 326

QY 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360

DB 327 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 386

QY 361 LRDEF 365

DB 387 LRDEF 391

RESULT 4

Q6EV66 PRELIMINARY; PRT; 391 AA.

ID Q6EV66

AC Q6EV66

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Protein-O-fucosyltransferase 1.

GN Name: fut12;

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovinae; Bos.

OC NCBI_TaxID=9913;

OX [1]

RN

QY		1	MPAGSWDPAGVLLYCPCMGREGNQADHFLGSLAFKLLNRTLAVPPWIEYOHHKKPPTNL	60
:		:	:::::	:
Db		20	LPAVSMDPAGVLLYCPCMGREGNQADHFGLSFAFKLLNRTLAVPPWIEVQHKKPPTNL	79
:~::~:		:	:::::	:
QY		61	HVSYYKYFKLBPQLAHRYISLEDFMFKLAPTHWPPEKRVAICYFEVAAQRSPDKTCKPK	120
:		:	:::::	:
Db		80	HVSYYKYFKLBPQLVTHRVISLEDFMFKLAPTHWPERRVAYCYFEVAAQRSDDKTCKMK	139
:~::~:		:	:::::	:
QY		121	EGNPGFPWDQFHYSFNKSSELFTCISFSASYRQWSORFSPKEHPVLALPGAQAOPPVLE	180
:		:	:::::	:
Db		140	EGNPGFPWDQFHYSFNKSSELFAGISFSASVKDQMIORFSPKEHPVLALPGAQAOPPVLE	199
:~::~:		:	:::::	:
QY		181	EHRPLQKVWMSDEWKVTGEAQIHAIHLVRVPVGVIHLIRIGSDWNACAMLKDGTAGSHFWA	240
:		:	:::::	:
Db		200	EHRPLQKVWMSDEWVRTGEAQIRAHLIRPVGVGHILRIGSDWNACAMLKDGTAHAHWA	259
:~::~:		:	:::::	:
QY		241	SPOCVGYSRSTAAPLTMTMCIPDLKEIQRAVKLVWRSLDAQSVVVATDSSESYPVELQQIF	300
:		:	:::::	:
Db		260	SPOCVGYSRHTAAPLTMTMCIPDLKEIRRALKLVWTALSQAQSVIVATDSSEYLPEIQQLF	319
:~::~:		:	:::::	:
QY		301	GKGKVVLSLKEVAQVDLYILQGADHFTIGNCVSSFTAfvkrddlqgrspssffgmdrrpp	359
:		:	:::::	:
Db		320	GKGKVVLSLKEVAQIDLYILQGADHFTIGNCVSSFTAfvrkreddlqcrisspfsgmdrrpp	378
:~::~:		:	:::::	:
RESULT 7				
OFUI MOUSE				
ID _OFUI MOUSE		STANDARD;	PRT; 393 AA.	
AC Q91ZB2;				
CD 28-FEB-2003 (Rel. 41, Created)				
DT 28-FEB-2003 (Rel. 41, Last sequence update)				
DT 25-OCT-2004 (Rel. 45, Last annotation update)				
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)				
DE (Peptide O-fucosyltransferase) (O-Fuct-1).				
DN Name=Pofuti;				
GN Mus musculus (Mouse).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OCC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX NCBI_TaxId=10090;				
RX [1]				
RN SEQUENCE FROM N.A.				
RP STRAIN=129/SvJ; TISSUE=Liver;				
RC MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;				
RX Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,				
RA Haliwanger R.S.;				
RA "Modification of epidermal growth factor-like repeats with O-fucose:				
RT molecular cloning and expression of a novel GDP-fucose protein O-				
RT fucosyltransferase."				
J Biol Chem. 276:40338-40345(2001).				
RL RT J Biol Chem. 276:40338-40345(2001).				

Db 18 AGSWDTAGYLLYPCMGRCFGNQAHEFLGALAFARALNRLTAVPPWIEYRHRPPYTNLHV 77
 QY 63 SVQYFKLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMKEG 122
 Db 78 PYEEYFKLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMKEG 137
 QY 123 NPFGPFDQFHVSNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLEEH 182
 Db 138 NPFGPFDQFHVSNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLEEH 197
 QY 183 RPLQKYMWSDEVMKVTGEGAIHAHLVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 242
 Db 198 RPLQKYMWSDEVMKVTGEGAIHAHLVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 257
 QY 243 QCVGYSRSTAAPLTMTCLPDLKEIORAVKLVWRSLEDAQSIVYVATDSSEYVPELQQLF 302
 Db 258 QCVGYSRSTAAPLTMTCLPDLKEIORAVKLVWRSLEDAQSIVYVATDSSEYVPELQQLF 317
 QY 303 KVKVSLKPEVAQVDLYLQADHFIQNCVSSFTAFVGRERDIQGRSPFFGMDRPPKLR 362
 Db 318 KIKVSLQPEVAQVDLYLQADHFIQNCVSSFTAFVGRERDIQGRSPFFGMDRPPKLR 377
 QY 363 DEF 365
 Db 378 DEF 380
 RESULT 10
 QY 07YRE6 PRELIMINARY; PRT; 351 AA.
 AC 07YRE6;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Protein O-fucosyltransferase 1b.
 OS Names: pofut1;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bos.
 NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lioriol C., Germot A., Dupuy F., Maftah A.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY344581; AAQ0233.1; -;
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 351 AA; 39576 MW; BP830F61A7296P42 CRC64;
 Query Match 80.1%; Score 1586; DB 2; Length 351;
 Best Local Similarity 88.6%; Pred. No. 4e-127;
 Matches 286; Conservative 20; Mismatches 9; Indels 8; Gaps 3;
 QY 1 MPAGSWDPAGVLLYPCMGRCFGNQAHEFLGALAFARALNRLTAVPPWIEYRHRPPYTNL 60
 Db 27 LPVGSWDPAGVLLYPCMGRCFGNQAHEFLGALAFARALNRLTAVPPWIEYRHRPPYTNL 86
 QY 61 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMK 120
 Db 87 HVSQYKFKLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMK 146
 QY 121 EGNPFGFDQFHVSNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLE 180
 Db 147 EGNPFGFDQFHVSNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLE 206
 QY 181 EHRPQKYMWSDEVMKVTGEGAIHAHLVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 240
 Db 207 EHRPQKYMWSDEVMKVTGEGAIHAHLVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 266
 QY 241 SPQCVGYSRSTAAPLTMTCLPDLKEIORAVKLVWRSLEDAQSIVYVATDSSEYVPELQQLF 300
 Db 267 SPQCVGYSRSTAAPLTMTCLPDLKEIORAVKLVWRSLEDAQSIVYVATDSSEYVPELQQLF 326

QY 301 KGVKVVSLK----PEVAQVDLYL 322
 Db 327 KGVK----SLSPFVPAKASEMSL--LLG 346
 RESULT 11
 QY 0640S0 PRELIMINARY; PRT; 380 AA.
 AC 0640S0;
 DT 25-OCT-2004 (TREMBlrel. 28, Created)
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue=Embryo;
 RA PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullah S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Tissue=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC082519; AAH82519.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 380 AA; 43570 MW; 8BB938CE776D5238 CRC64;
 Query Match 75.3%; Score 1490; DB 2; Length 380;
 Best Local Similarity 72.7%; Pred. No. 7.2e-119;
 Matches 261; Conservative 48; Mismatches 50; Indels 0; Gaps 0;
 QY 6 WDPAGVLLYPCMGRCFGNQAHEFLGALAFARALNRLTAVPPWIEYRHRPPYTNLHVSYQ 65
 Db 21 WDSGVLVYPCMGRCFGNQAHEFLGALAFARALNRLTAVPPWIEYRHRPPYTNLHVSYQ 80
 QY 66 KYFKLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMKEG 125
 Db 81 EFFQLEPLQAYHRVLSLEDFMEKLAATHMPPEKRVAYCEVAORSPDKKTCPMKEG 140
 QY 126 GPFWDFHVSFNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLEEH 185
 Db 141 GPFWDFHVSFNKSELFTGISFSASVREOWSRFSPEKHEPVLALPGAPAFQFVLEEH 200
 QY 186 QKYMWSDEVMKVTGEGAIHAHLVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 245
 Db 201 QRYVWSEKIVRAGEBQIQSLVVRPVYVGHILRIGSDWKNACAMLDKGTAGSHFMA 260
 QY 246 GYSRSTAAPLTMTCLPDLKEIORAVKLVWRSLEDAQSIVYVATDSSEYVPELQQLF 305

```
DB 261 GYDKYRAAPLTWEMCLPDLKEMRALTLWERSKARSVVIATDTSHTSEELOAFGLGDKVR 320
QY 306 VVSLKPEVAQVDLYLILGQADHFIIGNCVSSFTAFVKERDLQGRSPSSFGMDRPPKPLADE 364
DB 321 VVCLQPEVAQMDLYLILGQADHFIIGNCVSSFTAFVKERDLVHGRSPSSFGMDSPGAVQDE 379

RESULT 12
ID Q6EV71 PRELIMINARY; PRT; 396 AA.
AC Q6EV71;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Protein-O-fucosyltransferase.
GN Name=fut12;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Duncker I., Oriol R., Mollicone R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ781498; CSH03710.1;
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . . ; IEA.
DR InterPro; IPR000886; ER:target_S.
DR PROSITE; PS00014; ER:TARGET; UNKNOWN_1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 396 AA; 45056 MW; F90504E19507073F CRC64;

Query Match 75.3%; Score 1490; DB 2; Length 396;
Best Local Similarity 72.7%; Pred. No. 7.6e-119;
Matches 261; Conservative 48; Mismatches 50; Indels 0; Gaps 0;

QY 6 WDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLNRTLAVPPMIEYQHKKPPTNLHVSQ 65
DB 37 WDSGGYLLYPCPCMGFRGNQADHFLGSLAFKLNRTLAVPPMIVNHRPPYTNVHPYE 96
QY 66 KYKLEPLQAYHRVTSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKTCPMKEGNPF 125
DB 97 EFFQLEPLQAYHRVTSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKTCPMKEGNPF 156
QY 126 GPFWDQPHVSNKSELFTGISFSASYREQWSQRPSPKEHPVLALPGAPQAPVPLEEHRPL 185
DB 157 GPFWDHFNFTQSELDGTFISAYYKDVWNSRFPSPQHPFVIALPGAPQAPVPLEEHRSL 216
QY 186 QKYMVWSDENVKTGEAQIAHILVRPYVGIHLRIGSDWKACAMLDKGTAGSHFVMSPOCV 245
DB 217 QRYVWSEKIVRAGEBQIQSLVLRPYVGVHLRIGSDWKACAMLDKGTAGSHFVMSPOCV 276
QY 246 GYSESTAAPLTMTMCLPDLKEIQRVAVKLWRSLDQAQSVVATDSYVPELOQLFKGVK 305
DB 277 GYDKYRAAPLTWEMCLPDLKEMRALTLWERSKARSVVIATDTSHTSEELOAFGLGDKVR 336
QY 306 VVSLKPEVAQVDLYLILGQADHFIIGNCVSSFTAFVKERDLQGRSPSSFGMDRPPKPLADE 364
DB 337 VVCLQPEVAQMDLYLILGQADHFIIGNCVSSFTAFVKERDLVHGRSPSSFGMDSPGAVQDE 395

RESULT 13
QY7T028
ID Q7T028 PRELIMINARY; PRT; 395 AA.
AC Q7T028;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221).
GN Name=pofut1; Synonym=fut12;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Oriol R.;
RA Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ579536; CAE18459.1;
DR ZFIN; ZDB-GENE-040303-2; pofut1.
DR GO; GO:0046922; F:peptide-O-fucosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . . ; IEA.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 395 AA; 45165 MW; 15BBE06172542E8A CRC64;

Query Match 74.5%; Score 1474; DB 2; Length 395;
Best Local Similarity 73.9%; Pred. No. 1.8e-117;
Matches 263; Conservative 40; Mismatches 53; Indels 0; Gaps 0;

QY 5 SWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLNRTLAVPPMIEYQHKKPPTNLHVSQ 64
DB 33 TWDENGYLLYPCPCMGFRGNQADHFLGSLAFKLNRTLAVPPMIVVRRHSPPTNVHPY 92
QY 65 QKYKLEPLQAYHRVTSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKTCPMKEGNP 124
DB 93 SEYFOLEPLQAYHRVTSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKTCPMKEGNP 152
QY 125 GPFWDQPHVSNKSELFTGISFSASYREQWSQRPSPKEHPVLALPGAPQAPVPLEEHRP 184
DB 153 GPFWDHFGVDVDSVLFGLSFSSYQPHMKRPPFPKHPVLALPGAPQAPVPLEEHIG 212
QY 185 LQYVWSDENVKTGEAQIAHILVRPYVGIHLRIGSDWKACAMLDKGTAGSHFVMSPOC 244
DB 213 LOQFVWSDKIVQEGEGHRLNLRPYVGIHLRIGSDWKACAMLDKGTAGSHFVMSPOC 272
QY 245 GYSESTAAPLTMTMCLPDLKEIQRVAVKLWRSLDQAQSVVATDSYVPELOQLFKGVK 304
DB 273 VGYDQATALPLTMNCLPDLTEIRRAVKLWKNVKGTAQSVIATDSEHTTEIQKLPFGKV 332
QY 305 KVVSLKPEVAQVDLYLILGQADHFIIGNCVSSFTAFVKERDLQGRSPSSFGMDRPPK 360
DB 333 KVVSLQPDPAQVDLYLILGQADHFIIGNCVSSFTAFVKERDLVHGRSPSSFGMDRPPK 388

RESULT 14
QYAXS8 PRELIMINARY; PRT; 380 AA.
ID QYAXS8;
AC QYAXS8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein pofut1.
GN Name=pofut1;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Oriol R.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ514425; CAD55833.1;
DR InterPro; IPR000886; ER:target_S.
DR PROSITE; PS00014; ER:TARGET; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 380 AA; 43505 MW; C6B960E76D875293 CRC64;

Query Match 73.1%; Score 1447; DB 2; Length 380;
Best Local Similarity 70.6%; Pred. No. 3.4e-115;
Matches 255; Conservative 50; Mismatches 56; Indels 0; Gaps 0;

QY 4 GSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLNRTLAVPPMIEYQHKKPPTNLHVS 63
```

Db	19	GLYNSGYIVYCFMGRFGNQADHFLGSLAFKQVNRKTLVVPDWIVNHRPPYTNVHP	78
Qy	64	YQYFKLEPQAYHRVISLEDPMKELAPHTWPEKRVAYCFEVAQRSDDKKTCPMKEGN	123
Db	79	YEEFFQLEPLRQHQVISMEEFEQALPVLWPPTKLAVCFATAAQRSPDKSCPMKQGN	138
Qy	124	PGPFWDQFVSNKSELFTGISFSASYREWSQRFSPKEHPVLALPGAPAPFPVLEEHR	183
Db	139	PGPFWDHFDVDTQSELFDCITFSAYYKDVWISRPSPSHFVIALPGAPAPFPVLEEHR	198
Qy	184	PLQYVWSDVMVKTGBAQIAHILVRPYYGIHLRIGSDWKNCAMLKDGFTAGSHFNASPO	243
Db	199	SLQRYIVWSEKIVREGEQIRSLILVRPYYGIHLRIGSDWKDACKMLKDGFTAGSHVMASQA	258
Qy	244	CVGYSRSTAAPLTWMTCLPDLKEIQRAVKLWVRSLSAQSVYVATDSYVPELQQLFKGK	303
Db	259	CVCYDRYRAGPLTMDMCLPDLKEMRRALITLWVRSKARSYVIATDSPSYTAETQAFIGNE	318
Qy	304	KVVSLSKPEVAQVDLYILGOADHFIGNCVSSFTAFVKRERDLQGRPSSPFGMDRPPKLRD	363
Db	319	VRVVCLOPEVAQVDLYILGOADHFIGNCVSSFTAFVKRERDRVGRPSSPFGMDSPGVQD	378
Qy	364	E 364	
Db	379	E 379	
RESULT 15			
Q70AG7	070AG7	PRELIMINARY; PRT; 384 AA.	
AC	Q70AG7;	2004 (TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	Protein-O-fucosyltransferase.		
GN	Name=fut12;		
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Takifugu.		
OX	NCBI_TaxID=31033;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Martinez-Duncker I., Mollicone R., Candelier J.J., Oriol R.;		
RL	Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ606070; CAE54305.1; -.		
DR	GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.		
DR	InterPro; IPR000886; ER:target S.		
DR	PROSITE; PS00014; ER_TARGET; UNKNOWN_1.		
KW	Glycosyltransferase; Transferase.		
SQ	SEQUENCE 384 AA; 43650 MW; 7CAA6ADC90F40032 CRC64;		
Query Match 72.1%; Score 1427; DB 2; Length 384;			
Best Local Similarity 71.4%; Pred. No. 1.8e-113;			
Matches 255; Conservative 43; Mismatches 59; Indels 0; Gaps 0;			
Qy	6	WDPAGYLLYCPMGRFGNQADHFLGSLAFKLNRLTAVPPWIEYQHKKPPFTNLHVSQ	65
Db	23	WDQNGYVLYCPMGRFGNQADHFLGSLAFKLNRLTAVPPWIVYRHVPYTNVHIPYR	82
Qy	66	KYFKLEPQAYHRVISLEDPMKELAPHTWPEKRVAYCFEVAQRSDDKKTCPMKEGNPF	125
Db	83	DFQLEALSAYHRVVSLEDPMETLAPRYWPAGQRAYCFETAAQRTADKKSCPMKQGNPF	142
Qy	126	GPFWDFVSNKSELFTGISFSASYREWSQRFSPKEHPVLALPGAPAPFPVLEEHRPL	185
Db	143	GPFWDYVNVDFDESLVFGGIYFSAYYQPMWKKFPSPQHPVLPLPGAPAPFPVSEEHVL	202
Qy	186	QKYMVWSDVMVKTGBAQIAHILVRPYYGIHLRIGSDWKNCAMLKDGFTAGSHFNASPOCV	245
Db	203	QRYMVWSEKMWEEGDGHIKLLVRPYYGIHLRIGSDWKNCAMLKDGFTAGSHFNASPOCV	262

Qy	246	GYSRSTAAPLTWMTCLPDLKEIQRAVKLWVRSLSAQSVYVATDSYVPELQQLFKGKVK	305
Db	263	GYERQTALELTATMCLPDLGEILRAVKVWVKTSARSYVIATDSHSGDIEQLFNGKVK	322
Qy	306	VVSLKPEVAQVDLYILGOADHFIGNCVSSFTAFVKRERDLQGRPSSPFGMDRPPKLR	362
Db	323	VVSLRPELAQMDLYILGKADHFIGNCVSSFSAPVKRQDVQGLPSSFFGMDTFGNLR	379

Search completed: October 25, 2005, 15:35:23
Job time : 213.606 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:08:04 ; Search time 8270.58 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLPCMGK.....QGRPSFFGMDRPPKLRDF 365

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat_25102005_105432_6389/app_query.fasta_1.917
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.1.4968 @runat_25102005_105432_6389 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	1100	6	AR163457 Sequence
2	1979	100.0	1100	6	BD103196 O-fucosyl
3	1979	100.0	1167	6	CQ727777 Sequence
4	1979	100.0	1167	9	AJ781500 Pan trogl

ALIGNMENTS

RESULT 1	AR163457	1100 bp	DNA
LOCUS	Sequence 16 from patent US 6270987.		
DEFINITION	Sequence 16 from patent US 6270987.		
ACCESSION	AR163457		
VERSION	AR163457.1	GI:16234062	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1100)		
AUTHORS	Wang, Y. and Spellman, M. W.		
TITLE	O-fucosyltransferase		
JOURNAL	Patent: US 6270987-A 16 07-AUG-2001;		
FEATURES	Location/Qualifiers		
source	1..1100		
ORIGIN	/organism="unknown"		
	/mol_type="unassigned DNA"		

Alignment Scores:	8.91e-175	Length:	1100
Pred. No.:	1979.00	Matches:	365
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:			

5	1979	100.0	1300	6	AR163452	AR163452 Sequence
6	1979	100.0	1300	6	BD103189	BD103189 O-fucosyl
7	1979	100.0	1514	6	AR163451	AR163451 Sequence
8	1979	100.0	1514	6	BD103188	BD103188 O-fucosyl
9	1979	100.0	4560	9	HSW803330	AL812023 Homo sapi
10	1979	100.0	5189	9	DB0002	DB0002 Homo sapien
11	1979	100.0	5249	9	AF375884	AF375884 Homo sapi
12	1979	100.0	5266	6	AX780153	AX780153 Sequence
13	1979	100.0	11284	6	BD103190	BD103190 O-fucosyl
14	1979	100.0	11284	6	AX780153	AX780153 Sequence
15	1979	100.0	11284	6	AX780153	AX780153 Sequence
16	1979	100.0	11284	6	AX780153	AX780153 Sequence
17	1979	100.0	11284	6	AX780153	AX780153 Sequence
18	1979	100.0	11284	6	AX780153	AX780153 Sequence
19	1979	100.0	11284	6	AX780153	AX780153 Sequence
20	1979	100.0	11284	6	AX780153	AX780153 Sequence
21	1979	100.0	11284	6	AX780153	AX780153 Sequence
22	1979	100.0	11284	6	AX780153	AX780153 Sequence
23	1979	100.0	11284	6	AX780153	AX780153 Sequence
24	1979	100.0	11284	6	AX780153	AX780153 Sequence
25	1979	100.0	11284	6	AX780153	AX780153 Sequence
26	1979	100.0	11284	6	AX780153	AX780153 Sequence
27	1979	100.0	11284	6	AX780153	AX780153 Sequence
28	1979	100.0	11284	6	AX780153	AX780153 Sequence
29	1979	100.0	11284	6	AX780153	AX780153 Sequence
30	1979	100.0	11284	6	AX780153	AX780153 Sequence
31	1979	100.0	11284	6	AX780153	AX780153 Sequence
32	1979	100.0	11284	6	AX780153	AX780153 Sequence
33	1979	100.0	11284	6	AX780153	AX780153 Sequence
34	1979	100.0	11284	6	AX780153	AX780153 Sequence
35	1979	100.0	11284	6	AX780153	AX780153 Sequence
36	1979	100.0	11284	6	AX780153	AX780153 Sequence
37	1979	100.0	11284	6	AX780153	AX780153 Sequence
38	1979	100.0	11284	6	AX780153	AX780153 Sequence
39	1979	100.0	11284	6	AX780153	AX780153 Sequence
40	1979	100.0	11284	6	AX780153	AX780153 Sequence
41	1979	100.0	11284	6	AX780153	AX780153 Sequence
42	1979	100.0	11284	6	AX780153	AX780153 Sequence
43	1979	100.0	11284	6	AX780153	AX780153 Sequence
44	1979	100.0	11284	6	AX780153	AX780153 Sequence
45	1979	100.0	11284	6	AX780153	AX780153 Sequence

QY	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140
Db	361	GAAGGAAACCCCTTTGGCCCATTCCTGGATTCATGTTGATGTTTCAACAGTCGGAG	420
QY	141	LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer	160
Db	421	CTTTTACAGGCAATTCCTTCAGTCTCTACAGAGAACATGGAGCCAGAGATTTCT	480
QY	161	ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu	180
Db	481	CCAAAGGAACATCCGGTCTTGGCTCCAGAGAGCCAGCCAGTTCCTCCCTAGAA	540
QY	181	GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu	200
Db	541	GAACACAGGCCACTACAGAGATGATGATGATGATGATGATGATGATGATGATGAT	600
QY	201	AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer	220
Db	601	GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGCATTCATCTGGCATTTGCT	660
QY	221	AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
Db	661	GACTGGAAGAACCTGTCGATGCTGAAGAGCGGACTGCGAGCTCGCACTTCATGGCC	720
QY	241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
Db	721	TCTCCGAGTGTGGGCTACAGCCGAGCAGCAGCCGCCCTTCACGATGATATGTGC	780
QY	261	LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla	280
Db	781	CTGCCTGACCTGAAGAGATCCAGAGGCTGTGAAGTCTGGGTGAGTCTGCTGGATGCC	840
QY	281	GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe	300
Db	841	CAGTCGGTCTACGTGTGCTACTGATTCGAGATGATGCTGCTGCTGCTGCTGCTGCT	900
QY	301	LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle	320
Db	901	AAAGGAAGTGAAGTGGTGGAGCTGAAGCTGAGTGGCCAGGTCGACCTGTATCATC	960
QY	321	LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys	340
Db	961	CTCGCCCAAGCCGACCACTTATTTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCT	1020
QY	341	ArgGluArgAspLeuGlnGlyArgProSerPhePheGlyMetAspArgProProLys	360
Db	1021	CGGAGCGGAGCTCCAGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1080
QY	361	LeuArgAspGluPhe	365
Db	1081	CTCGGAGCAGTTTC	1095
RESULT 3			
CQ727777			
LOCUS	CQ727777 1167 bp DNA linear PAT 03-FEB-2004		
DEFINITION	Sequence 13711 from Patent WO02068579.		
ACCESSION	CQ727777		
VERSION	CQ727777.1 GI:42294771		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.		
TITLE	Kits, such as nucleic acid arrays, comprising a majority of human exons or transcripts, for detecting expression and other uses thereof		
JOURNAL	Patent: WO 02068579-A 13711 06-SEP-2002;		
FEATURES	PE Corporation (NY) (US)		
source	Location/Qualifiers		
	1..1167		

/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Alignment Scores:			
Pred. No.:	9.63e-175	Length:	1167
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-09-774-954-2 (1-365) x CQ727777 (1-1167)			
QY	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
Db	70	ATGCCCGCGGCTCCCTGGGACCCGCGGTTACCTGCTCTTACTGCCCTGCGATGGGGCGC	129
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
Db	130	TTTGGGAACAGGCCGATCACCTTCTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT	189
QY	41	ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu	60
Db	190	ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAAAGCTCTTTTCAACCACTC	249
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle	80
Db	250	CATGTCTCTACAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC	309
QY	81	SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal	100
Db	310	AGCTTGGAGGATTCATGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGTG	369
QY	101	AlaTyrCysPheGluValAlaGlnArgSerProAspLysLysThrCysProMetLys	120
Db	370	GCATATCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATAAGAGAGCTGCCCATGAAG	429
QY	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140
Db	430	GAAGGAAACCCCTTTGGCCCATTCCTGGGATTCAGTTCATGTGAGTTTCAACAAGTCGG	489
QY	141	LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer	160
Db	490	CTTTTACAGGCAATTCCTTCAGTCTCTACAGAGAACATGGAGCCAGAGATTTCT	549
QY	161	ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu	180
Db	550	CCAAAGGAACATCCGGTGTCTTGGCTTGGCCCTGCCAGGAGCCAGCCAGTTCCTCTAG	609
QY	181	GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu	200
Db	610	GAACACAGGCCACTACAGAGATGATGATGATGATGATGATGATGATGATGATGAT	669
QY	201	AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer	220
Db	670	GCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCCATTTGGCTCT	729
QY	221	AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
Db	730	GACTGGAAGAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	789
QY	241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
Db	790	TCTCCGAGTGTGGGCTACAGCCGAGCAGCAGCCGCCCTTCACGATGATATGTGC	849
QY	261	LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla	280
Db	850	CTGCCTGACCTGAAGAGATCCAGAGGCTGTGAGCTCTGGGTGAGGTCCGCTGATGCC	909
QY	281	GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe	300

Db	1030	CTCGGCCAAGCTGACCACTTTATTGGCAA	CTGTGTCCTCCTCTCCATCTGCGTGTGAAG	1089
Qy	341	ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys	360	
Db	1090	CGGAGCGGACCTCCAGGGAGCGCGTCTCTTTCTTCGGCATGGACAGGCCCCCTTAAG	1149	
Qy	361	LeuArgAspGluPhe	365	
Db	1150	CTGGGAGCAGTTTC	1164	
 RESULT 5 ARI163452				
LOCUS	ARI163452	Sequence 6 from patent US 6270987.	1300 bp	DNA linear PAT 17-OCT-2001
DEFINITION	ARI163452			
ACCESSION	ARI163452			
VERSION	ARI163452.1	GI:16234057		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 1300)			
AUTHORS	Wang,Y. and Spellman,M.W.			
TITLE	O-fucosyltransferase			
JOURNAL	Patent: US 6270987-A 6 07-AUG-2001;			
FEATURES	Location/Qualifiers			
source	1..1300			
	/organism="unknown"			
	/mol_type="unassigned DNA"			
 ORIGIN				
Alignment Scores:				
Pred. No.:	1.11e-174	Length:	1300	
Score:	1979.00	Matches:	365	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	6	Gaps:	0	
 US-09-774-954-2 (1-365) x ARI163452 (1-1300)				
Qy	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20	
Db	136	ATGCCCGGGCTCTCTGGACCCGCGGTACTGTCTACTGCCCTGCATGGGCGC	195	
Qy	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40	
Db	196	TTTGGGAACACGAGCGCATCACTTCTTGCGCTCTCTGGCATTTGCCAAAGCTGCTAAACCGT	255	
Qy	41	ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProPheThrAsnLeu	60	
Db	256	ACCTTGGCTGTCCTCTTGATTGAGTACCAGCATCACAGCTCTCTTTTCCCACACCTC	315	
Qy	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValile	80	
Db	316	CATGTGTCCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC	375	
Qy	81	SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal	100	
Db	376	AGCTTGGAGGATTCATGGAGAAGCTGGCACCCACCTGCGCCCTTGGAGCGGGTG	435	
Qy	101	AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys	120	
Db	436	GCATACTCTTTGAGTGGGAGCCCGAGGAGCCAGATAGAAGACGTGCCCATGAAG	495	
Qy	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140	
Db	496	GAAGAAACCCCTTTGGGCCAATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	555	
Qy	141	LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnInTrpSerGlnArgPheSer	160	
Db	556	CTTTTTACAGGCAATTCCTTCAGTGCTTCCTACAGAGAACAAATGGAGCCAGAGATTTTCT	615	

121 GluGlyAsnProPheGlyProPheTyrAspGlnPheHisValSerPheAsnLysSerGlu 140
 361 GAAGGAAACCCCTTTGGCCATCTCGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420
 141 LeuPheThrGlyLeuSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
 421 CTTTTCACAGGATTCCTTCAGTGTCTCTACAGAGAACATGGAGCCAGAGATTTTCT 480
 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
 481 CCAAGGAACATCCGGTGTCTGCGCCAGGAGCCAGCCAGTTCCTCCGCTCTAGAA 540
 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 541 GAACACAGCCCACTACAGAGTACATGTGTGTGAGAGCAATGTGAGAGCGGAGAG 600
 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 601 GCCCAGATTTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGGCATTTGCTCT 660
 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 661 GACTGGAAGAACCCCTGTGCGATGCTGAAGAGCGGACTGCAGGCTCGCAGCTTCATGGCC 720
 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
 721 TCTCCGAGTGTGGGCTACACCCGAGCAGCAGCGGCCCTCCAGTACATGATGTGTC 780
 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
 781 CTGCCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGGTGGTGC 840
 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTTC 900
 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
 901 AAAGGAAGGTGAAGGTGGAGCCGGAAGCCCTGAGGTGGCCAGGTTCGACCTGTACATC 960
 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
 961 CTGCGCCACCGCCACCTTATTTGCAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 341 ArgGluArgAspLeuGlnArgProSerSerPhePheGlyMetAspArgProLys 360
 1021 CCGGAGCGGACCTCAGGCGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
 361 LeuArgAspGluPhe 365
 1081 CTGCGGAGCAGATTTC 1095

RESULT 8
 BD103188
 LOCUS BD103188 1514 bp DNA linear PAT 27-AUG-2002
 DEFINITION O-fucosyltransferase.
 ACCESSION BD103188
 VERSION BD103188.1 GI:22648762
 KEYWORDS JP 2001527389-A/1.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 1514)
 AUTHORS Wang, Y. and Spellman, M.W.
 TITLE O-fucosyltransferase
 JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;
 GENENTECH INC
 COMMENT OS Unidentified
 PN JP 2001527389-A/1
 PD 25-DEC-2001
 PF 17-DEC-1997 JP 1998532877
 PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
 YANG WANG, MICHAEL W SPELLMAN

PC C12N15/54, C12N9/10, C07K16/40
 CC Strandedness: Single;
 CC Topology: Linear;
 CC O-fucosyltransferase
 FH Key Location/Qualifiers
 FT source 1..1514
 FT /organism='Unidentified'.
 FEATURES
 source
 1..1514
 /organism='unidentified'
 /mol_type='genomic DNA'
 /db_xref='taxon:32644'
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.35e-174 Length: 1514
 Score: 1979.00 Matches: 365
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-774-954-2 (1-365) x BD103188 (1-1514)
 QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 DB 1 ATGCCCGGGGCTCTCGGACCGCGCGGTACTCTCTACTGCCCCCTGCATGGGGCGC 60
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArg 40
 DB 61 TTTGGGAACAGCGCGCATCACTTCTTGGGCTCTCTGGCATTTTGCAGAGCTGTAAACCGT 120
 QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysHisLysProProPheThrAsnLeu 60
 DB 121 ACCTTGGCTGTCTCTCTCTTGGATTGATACCATCATCAGGCTCTCTTTCACCACTC 180
 QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
 DB 181 CATGTCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
 QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
 DB 241 AGCTTGGAGATTTCATGGAGAAGCTGGCACCCACCTCTGGCCCCCTGAGAGCGGGTG 300
 QY 101 AlaTyrCysPheGluValAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 DB 301 GCATATCTGCTTTGAGTGGCAGCCAGAGCCAGATAGAGAGCTGCCCATCAAG 360
 QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
 DB 361 GAAGGAACCCCTTTGGCCATCTCGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 420
 QY 141 LeuPheThrGlyLeuSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
 DB 421 CTTTTCACAGGATTCCTTCAGTGTCTCTACAGAGAACATGGAGCCAGAGATTTTCT 480
 QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
 DB 481 CCAAGGAACATCCGGTGTCTGCGCCAGGAGCCAGCCAGTTCCTCCGCTCTAGAA 540
 QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 DB 541 GAACACAGCCCACTACAGAGTACATGTGTGTGAGAGCAATGTGAGAGCGGAGAG 600
 QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 DB 601 GCCCAGATTTCGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGGCATTTGCTCT 660
 QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 DB 661 GACTGGAAGAACCCCTGTGCGATGCTGAAGAGCGGACTGCAGGCTCGCAGCTTCATGGCC 720
 QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260

Db 721 TCTCCGACGTGTGGGTACAGCCGACAGCGGCGCCCTCACGATGACTATGTGC 780
Qy 261 LeuProAspLeuGluValLeuGlnAlaValLysLeuTyrValArgSerLeuAspAla 280
Db 781 CTGCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGTGCCTGGATGCC 840
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 841 CAGTCGGTCTAGTTGCTACTGATTCGAGAGTATATGCGCTGAGCTCCACAGACTCTTC 900
Qy 301 LysGlyValValValValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 901 AAGGGAAGGTGAAGTGGTGGAGCTGAAGCTGAAGTGGAGTGGAGTGGAGTGGAGTGG 960
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 961 CTGGGCAAGCGACCACTTATTGGCACTGTGTCTCTCTCACTGCTTTTGTGAAG 1020
Qy 341 ArgGluArgAspLeuGlnGlyValGlyProSerPhePheGlyMetAspArgProProLys 360
Db 1021 CGGGAGCGGGAGCTCCAGGGAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Qy 361 LeuArgAspGluPhe 365
Db 1081 CTGGGGAGCAGGTTTC 1095

RESULT 9
HSM803330 4560 bp mRNA linear PRI 13-MAY-2003
LOCUS Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
DEFINITION AL832023
ACCESSION AL832023.1 GI:21732563
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4560)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,W. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp451J114) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
FEATURES
Location/Qualifiers
1..4560
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp451J114"
/db_xref="taxon:9606"
/clone="DKFZp451J114"
/tissue_type="human skeletal muscle"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host
DH103; sites NotI + SalI"
/dev_stage="adult"
4538

polyA_site
ORIGIN

Alignment Scores:
Pred. No.: 5,66e-174 Length: 4560
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0
US-09-774-954-2 (1-365) x HSM803330 (1-4560)
Qy 1 MetProAlaGlySerTyrAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 51 ATGCTGCGGGCTCTGCGGACCCGGCGGTTCCTCTCTACTGCCCCCTGCATGGGGCGC 110
Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArg 40
Db 111 TTTGGGAAACAGGCCGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 170
Qy 41 ThrLeuAlaValProProTyrIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
Db 171 ACCTTGGCTGTCTCTCTGATGAGTACCAACATCAAGCCCTCTTCCACCACTC 230
Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db 231 CATGTGTCTCTACAGAACTTCTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 290
Qy 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTyrProGluLysArgVal 100
Db 291 AGCTTGGAGATTTTCATGGAGAGCTGGCACCCACCCACTGGCCCCCTGAGAGCGGTG 350
Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 351 GCATATCTGCTTTGAGTGGAGCCCGAGGAGCCAGATGAAGACGTGCCCTCTGAAG 410
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTyrSerGlnArgPheSer 160
Db 471 CTTTTTACAGGCATTTCTCTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 530
Qy 161 ProlGluHisProValLeuAlaLeuProGlyValAlaGlnPheProValLeuGlu 180
Db 531 CCAAGAGACATCCGGTGTCTTGGCTTGCAGAGAGCCCGAGCCCACTTCCCGCTCTAGAG 590
Qy 181 GluHisArgProLeuGlnLysTyrMetValTyrSerAspGluMetValLysThrGlyGlu 200
Db 591 GAACACAGGCCACTACAGAGTACATGATGTATGTTCAGAGAAATGTGAACGGAGAG 650
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 651 GCCCAGATTTCATGCCCACTTGTCCGGCCCTATGTGGCATTTTCATCTGCCATTTGGCTCT 710
Qy 221 AspTyrLysAlaAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 711 GACTGGAGAACGCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTTCATGGCC 770
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 771 TCTCCGAGTGTGGGCTACAGCCCGAGCAGCCGCGCCCTCACGATGCTATGTGC 830
Qy 261 LeuProAspLeuLysGluIleGlnAlaValLysLeuTyrValArgSerLeuAspAla 280
Db 831 CTGCTGACCTGAAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCTCGATGCC 890
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 891 CAGTCGGTCTAGTTGCTACCGATTCCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC 950
Qy 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 951 AAGGGAAGGTGAAGTGGTGGAGCTGAAGCTTGAAGTGGGCGCCAGGTCGACCTGTACATC 1010
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 1011 CTGGGCCAAGCGGACCACTTATTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1070
Qy 341 ArgGluArgAspLeuGlnGlyArgProSerPhePheGlyMetAspArgProProLys 360

Db 1071 CGGAGGGGACCTCCAGGGGAGCGCTCTTCTTCTTGGCATGACAGGCCCTTAAG 1130
QY 361 LeuArgAspGluPhe 365
Db 1131 CTGCGGGACGAGTTC 1145
RESULT 10
D80002 5189 bp mRNA linear PRI 10-JAN-2003
LOCUS Homo sapiens KIAA0180 mRNA, complete cds.
DEFINITION D80002
ACCESSION D80002.2 GI:20521837
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Nagase, T., Seki, N., Iahikawa, K., Tanaka, A. and Nomura, N.
TITLES Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1
JOURNAL DNA Res. 3 (1), 17-24 (1996)
MEDLINE 96281124
PUBMED 8724849
REFERENCE 2
AUTHORS Chiang, P.W., Wang, S., Smithivas, P., Song, W.J., Ramamoorthy, S., Hillman, J., Puett, S., Van Keuren, M.L., Crombez, E., Kumar, A., Glover, T.W., Miller, D.E., Tsai, C.H., Blackburn, C.C., Chen, X.N., Sun, Z., Cheng, J.F., Korenberg, J.R. and Kurnit, D.M.
TITLES Identification and analysis of the human and murine putative chromatin structure regulator SUPT6H and Supt6h
JOURNAL Genomics 34 (3), 328-333 (1996)
MEDLINE 96374824
PUBMED 8786132
REFERENCE 3 (bases 1 to 5189)
AUTHORS Ohara, O., Negase, T., Kikuno, R. and Nomura, N.
TITLES Direct Submission
JOURNAL Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913)
COMMENT On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES
Location/Qualifiers
1..5189
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hs0256781"
/sex="male"
/cell_line="KG-1"
/cell_type="myeloblast"
/tissue_type="brain"
/clone_id="p1uescriptII SK plus"
/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1066 was derived from pf09933 and 1067 - 5189 was derived from ha02567)."
1..5189
/gene="KIAA0180"
1..1213
/gene="KIAA0180"
/citation=[2]
/codon_start=2
/protein_id="BA011497.2"
/db_xref="GI:20521838"
/translation="PPRLCAAGSGRADMGAAWRPLSVSFLLLLPLPMPAGSW DPAGLLYPCWGRFGNQADHFLGSLAPAKLLNTLAVPPWIEYQHKPPTNLTNLSY QYKLEPLQAYHVRVLEDFMEKLAHPHPPVAYCFEVAQAQSPDFKRTCPMKGG NPFPGFMDQFVMSKSELTFTGISFSASYEQNSQFSPKHPVIALPAGAPQPVLE EHRPLOCKVMSDMWKTGEAQIHAIHVRPVYVGHIRIGSDWKNAKMLKDGTAGSHF MASPOCVGYSRSTAPLTMTMCLPDLKEIOPAVKLVRSILDAQSVVATDSYVPEL QQLFAGKVKVSLRFEVAQVDLYILGQADHFIQNCVSVSFTAFVKEERDLQRPSSFFG

ORIGIN MDRPPKLRDEF"
Alignment Scores:
Pred. No.: 6,7e-174 Length: 5189
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-774-954-2 (1-365) x D80002 (1-5189)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 116 ATGCTTGGGGTCTCTGGGACCGCGGTTACTGCTCTACTGCGCTCATGTGGGCGC 175
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArg 40
Db 176 TTTGGGAACCCAGCCGATCATCTTTGGGCTCTCTGGCATTTTCAAGAGCTGCTAAACCGT 235
QY 41 ThrLeuAlaValProTrpTlleGluTyrGlnHisLysProPheThrAsnLeu 60
Db 236 ACCTTGCTGCTCCTCTCTGGATTGATACCATCACAAGGCTCTTTTCAACCACTC 295
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db 296 CATGTGCTCTACAGAACTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 355
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
Db 356 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTGGGCCCCCTGAGAAGCGGGTG 415
QY 101 AlaTyrCysPheGluValAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 416 GCATACTGCTTTGAGGTGGCAGCCAGAGCCAGATAGAGAGAGCTGCCCATGAG 475
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 476 GAAGGAACCCCTTTGGCCATTTCTGGGATCAGTTTCATGTGATTTCAACAAGTCGGAG 535
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 536 CTTTTCACAGGCATTTCTTTCAGTGTCTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 595
QY 161 ProLysGluHisProValLeuAlaLeuProGlyValaProAlaGlnPheProValLeuGlu 180
Db 596 CCAAGGAACATCCGGTGTCTGCCCTGCCAGGAGCCCGCCAGTTCCTCCGCTCTAGAG 655
QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 656 GAACACAGGCCCATCTACAGAGATACATGGTATGGTCAGACGAAATGGTGAAGACGGAGAG 715
QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 716 GCCCAGATTTCATGCCACCTTGTCCGCCCTATGTGGGCATTTCTCTGCGCATTTGGTCT 775
QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 776 GACTGGAAGAACGCTGTGCCATGCTGAAGGACGGGACTGAGGCTCGACTTCATGCCC 835
QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrCys 260
Db 836 TCTCCGAGTGTGTGGGCTACAGCCGAGCAGCAGCGCCCCCTCCAGTATGTATGTC 895
QY 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db 896 CTGCTTGCCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGCGATGCC 955
QY 281 GlnSerValTyrValAlaThrAspSerGlnSerTyrValProGluLeuGlnLeuPhe 300
Db 956 CAGTCCGCTACAGTGTGCTACTGATTCCGAGAGTTATGTGCTGAGCTTCCAAAGCTCTTC 1015


```
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FH Key 1..11284 Location/Qualifiers
FT source 1..11284 /organism='Unidentified'.
FEATURES
    source
        Location/Qualifiers
            1..11284 /organism='Unidentified'
            /mol_type='genomic DNA'
            /db_xref='taxon:32644'

ORIGIN
Alignment Scores:
Pred. No.: 1..846-173 Length: 11284
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Dbs: 6 Gaps: 0

US-09-774-954-2 (1-365) x BD103190 (1-11284)
QY 1 MetProAlaGlySerThrPaspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 4236 ATGCCCGGGGCTCTCTGGGACCGCGCGGTACCTCTCTACTGCCCTGCATGGGGCC 4295
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 4296 TTTGGGAACCGCGCGATCACTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 4355
QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
DB 4356 ACCTTGGCTGTCCCTTGGATTGAGTACAGCATCAAGAGCTCTCTTTCACCAACCTC 4415
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
DB 4416 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATTGGGTCAATC 4475
QY 81 SerLeuLysPheMetGlyLysLeuAlaProThrHisTyrProProGluLysArgVal 100
DB 4476 AGCTTGGAGATTTCTGGAGAGCTGGACCCACCCACTGCCCCCTTGAGAGCGGGT 4535
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
DB 4536 GCATCTGCTTTGAGTGGGAGCCAGCAGAGCCAGATAGAGAGCTGCCCATGAA 4595
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
DB 4596 GAAGGAACCCCTTTGGCCCATCTGGGATCATGTGTGAGTTTCAACAAAGTCGGAG 4655
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
DB 4656 CTTTTTACAGGCATTTCTCTCAGTGTCTTCTACAGAGAACATGAGCCAGAGATTTTCT 4715
QY 161 ProLysGluHisProValLeuAlaLeuProGlyValaProAlaGlnPheProValLeuGlu 180
DB 4716 CCAAGGAACATCCGGTCTTGGCTTCCAGGAGCCCGAGCCAGCTTCCCGTCTTAGAG 4775
QY 181 GluHisArgProGluLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
DB 4776 GAACACAGGCCCATACAGAGTACATGATATGTCTGAGACGAATGTTGAGACGGGAGAG 4835
QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgTrpGlySer 220
DB 4836 GCCAGATTCATGCCCATCTTGTCCGCCCTATGTGGCATTCATCTGCGCATTTGCTCT 4895
QY 221 AspTrpLysAlaAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
DB 4896 GACTGAAGAAGCGCTGTGCCATCTCAAGACGGGACTGCAAGCTTCATGCGCC 4955
QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
```

```
DB 4956 TCTCCGAGTGTGGCTACAGCCGAGCAGCGGGCCCCCTCAGCATGACTATGTGC 5015
QY 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
DB 5016 CTGCTGAGCTGAAGGAGATCCAGAGGCTGTGAGACTCTGGGTGAGTCTCTGGATGCC 5075
QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
DB 5076 CAGTCGGTCTAGCTTGGCTACTGATTCGAGAGTATTATGTCCTGAGCTCCAACAGCTTTC 5135
QY 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
DB 5136 AAAGGGAAGCTGAAGGTGGTGAAGCTGAAGCTGAGCTGGCCAGGTGCAGCTGTACATC 5195
QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
DB 5196 CTGGCCAAAGCGACCACTTTATTGGCAACTGTGTCTCTCTCACTGCTTTGTGAAG 5255
QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArgProProLys 360
DB 5256 CGGAGGGGAGACTCCAGGGAGGCGCTCTTCTTCTTGGGATGACAGAGCCCTTAAG 5315
QY 361 LeuArgAspGluPhe 365
DB 5316 CTGGGAGCAGATTC 5330

RESULT 14
LOCUS AY344580
DEFINITION Bos taurus protein O-fucosyltransferase 1a (pofut1) mRNA, complete cds.
ACCESSION AY344580
VERSION AY344580.1 GI:33303529
KEYWORDS .
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE 1 (bases 1 to 1176)
AUTHORS Loricl C., Germot A., Dupuy F. and Maftah, A.
TITLE Genomic organization and expression profile of O-fucosyltransferase
genes, pofut1 and pofut2, in Bos taurus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1176)
AUTHORS Loricl, C.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie,
Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges,
87060, France
FEATURES
    source
        Location/Qualifiers
            1..1176 /organism="Bos taurus"
            /mol_type="mRNA"
            /db_xref="taxon:9913"
            1..1176 /gene="pofut1"
            1..1176 /gene="pofut1"
            /note="O-Fuct-1a"
            /codon_start=1
            /product="protein O-fucosyltransferase 1a"
            /protein_id="AA02332.1"
            /db_xref="GI:33303530"
            /translation="MGAAWAPSLPPRVSLALLPLPLPGVSWDPACGLLYCPCM
            GFGNOADHFPGSLAFAPAKLNLRTLAVPMWIEYOHHPPTTNVHVSYOKYFKLPIQAY
            HAVSLDHPFMKLAFTWHPPEKAVAYCAVAAQSPDKTCPMKEGNPFGPFDQFHV
            SNKSELFAQISFSASYKDWIQRFSPBEHFWALPGAPAQFPVLEHRLPKYMWYS
            DEWVRIGQIHAHLIRPYVGIHLIRIGSDWKACAMLKADGTAGAHFNASPCQCVSEH
            TAPATWTMCLPDLKGIIRRAKLVWITALNQSVYIATDSRSLFEIQLPFGVKVKTYS
            LKPEVAQIDLVLQADHFIGNCVSSFTAFVKRKLHGRFSFFGMDRPPQRDBF"
    ORIGIN
```

321 Qy LeuGlyGlnAlaAspHisPheIleGlyAsnCyveAlserSerPheThrAlaPheValIlys 340
 1039 Db CTCGGCCAGCTGACACATTTATTGGCAACTGTGTCTCTCTCTCTCACTGCTCTGTTAAG 1098

341 Qy ArGGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
 1099 Db CGGGAACGGACCTCCATGGAGACCATCTCTCTTTGGCATGGACAGGCCCCCTCAG 1158

361 Qy LeuArgAspGluPhe 365
 1159 Db CTGCGGGATGAATC 1173

RESULT 15
 AJ781503
 LOCUS Bos taurus mRNA for protein-O-fucosyltransferase 1 (fut12 gene).
 DEFINITION
 AJ781503 1176 bp mRNA linear MAM 07-JUL-2004
 ACCESSION
 AJ781503.1 GI:50057075
 VERSION fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
 KEYWORDS Bos taurus (cow)
 SOURCE Bos taurus
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.

REFERENCE 1
 AUTHORS Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and Oriol,R.
 TITLE A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs
 JOURNAL Glycobiology 13 (12), 1C-5C (2003)
 PUBMED 12966037

REFERENCE 2
 AUTHORS Martinez-Duncker,I., Oriol,R. and Mollicone,R.
 TITLE Phylogeny of fucosyltransferases
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1176)
 AUTHORS Oriol,R.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE

FEATURES
 source Location/Qualifiers
 1..1176
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 1..1176
 /gene="fut12"
 1..1176
 /gene="fut12"
 /function="adds fucose to Ser or Thr"
 /note="pofut3"
 /codon_start=1
 /product="protein-O-fucosyltransferase 1"
 /protein_id="CAH03714.1"
 /db_xref="GI:50057076"
 /translations="MGAAWAPSLPPRVSLILLILLPLGLPVGSMNPAGYLLFCPCNKH
 GRFGNADHPGLSIAFAPLMNRITLAVPPWIEYQHKPFTTNVHVSQKYFKLEPQAA
 HVSLEDFNEKLAPTHPEPEKAVYCFEVAORSPDKTCKPMKEGMPFGFWDQFHV
 SFNKISLEAGISFSASTQWIRFSPSEHPVLALPCDAQFPVLEHRPLQKYMVAW
 DENVRITCGAIHAHLIRPYVGIHLRIGSDWKNACAMLDGTAGAHFWASPCVGVSRIR
 TTAPLMTGCLPRLKALRMTALNAQSVYIATDSSELYLPEIQQLFPGKVKVWS
 LKPEVAQIDLYILQADHFICNCVSVFTAFVKRERDLHRPSSFGMDRPLQRLDEF

ORIGIN

Alignment Scores:
 Pred. No.: 2.52e-165 Length: 1176
 Score: 1878.00 Matches: 337
 Percent Similarity: 97.81% Conservative: 20
 Best Local Similarity: 92.33% Mismatches: 8
 Query Match: 94.90% Indels: 0
 DB: 4 Gaps: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 214.353 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLCPMGR.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10G_PUBCOMB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	100.0	365	11	US-09-774-954-2
2	1979	100.0	388	14	US-10-301-822-169
3	1979	100.0	397	11	US-09-774-954-6
4	1844	93.2	343	11	US-09-774-954-17
5	831	42.0	402	20	US-11-097-143-18162
6	809	40.9	165	11	US-09-833-245-366
7	541.5	27.4	474	11	US-09-774-954-8
8	353	17.8	61	11	US-09-774-954-9
9	295	14.9	61	11	US-09-774-954-3
10	164.5	8.3	328	16	US-10-471-450-26
11	159.5	8.1	490	20	US-11-097-143-36192
					Sequence 2, Appli
					Sequence 169, App
					Sequence 6, Appli
					Sequence 17, Appl
					Sequence 18162, A
					Sequence 166, App
					Sequence 8, Appli
					Sequence 9, Appli
					Sequence 3, Appli
					Sequence 26, Appl
					Sequence 36192, A

12	142	7.2	429	16	US-10-471-450-8	Sequence 8, Appli
13	120	6.1	150	18	US-10-820-474A-102	Sequence 102, App
14	117	5.9	543	15	US-10-424-599-145078	Sequence 145078,
15	106.5	5.4	441	16	US-10-437-963-162121	Sequence 162121,
16	103	5.2	561	16	US-10-425-115-328476	Sequence 328476,
17	103	5.2	581	15	US-10-425-114-65103	Sequence 65103, A
18	102.5	5.2	260	14	US-10-128-714-3312	Sequence 3312, Ap
19	101	5.1	28	11	US-09-774-954-14	Sequence 14, Appl
20	101	5.1	538	16	US-10-437-963-113919	Sequence 113919,
21	100	5.1	207	15	US-10-309-290-120	Sequence 120, App
22	100	5.1	207	15	US-10-309-290-122	Sequence 122, App
23	98.5	5.0	446	15	US-10-425-114-68801	Sequence 68801, A
24	95.5	4.8	442	16	US-10-425-115-237127	Sequence 237127,
25	95	4.8	385	14	US-10-156-761-13301	Sequence 13301, A
26	94.5	4.8	4773	15	US-10-287-226-332	Sequence 322, App
27	93	4.7	713	16	US-10-437-963-170855	Sequence 170855,
28	93	4.7	1374	16	US-10-437-963-132782	Sequence 132782,
29	92.5	4.7	1251	15	US-10-282-122A-63230	Sequence 63230, A
30	92	4.6	348	15	US-10-282-122A-53200	Sequence 53200, A
31	92	4.6	650	15	US-10-282-122A-73270	Sequence 73270, A
32	91	4.6	1381	14	US-10-032-585-7784	Sequence 7784, Ap
33	91	4.6	1381	16	US-10-451-467A-426	Sequence 426, App
34	90.5	4.6	353	15	US-10-309-290-118	Sequence 118, App
35	90.5	4.6	455	18	US-10-450-763-55399	Sequence 55399, A
36	90.5	4.6	568	15	US-10-282-122A-71715	Sequence 71715, A
37	90	4.5	1205	16	US-10-437-963-161150	Sequence 161150,
38	90	4.5	1315	18	US-10-450-763-40320	Sequence 40320, A
39	89.5	4.5	218	17	US-10-732-923-10961	Sequence 10961, A
40	89	4.5	511	16	US-10-425-115-324710	Sequence 324710,
41	89	4.5	518	15	US-10-425-114-53636	Sequence 53636, A
42	89	4.5	551	16	US-10-425-115-357766	Sequence 357766,
43	89	4.5	581	15	US-10-425-114-59580	Sequence 59580, A
44	88.5	4.5	314	16	US-10-425-115-283759	Sequence 283759,
45	88.5	4.5	991	16	US-10-437-963-181268	Sequence 181268,

ALIGNMENTS

RESULT 1

US-09-774-954-2
; Sequence 2, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-774-954-2

Query Match          100.0%; Score 1979; DB 11; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.1e-195;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYCPCMGFRGNQADHFLGSLAFAPKLLNRTLAVPPWIEYQHKKPPTNL 60
DB 1 MPAGSWDPAGYLLYCPCMGFRGNQADHFLGSLAFAPKLLNRTLAVPPWIEYQHKKPPTNL 60
QY 61 HVSQYKFKLEPLQAYHRVVISLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKKTCPMK 120
DB 61 HVSQYKFKLEPLQAYHRVVISLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKKTCPMK 120
QY 121 EGNPFGPFWDQFHVSFNKSELFTGTSFSASYRQWSQRFSPKEHPVLALPGAPQFPVLE 180
DB 121 EGNPFGPFWDQFHVSFNKSELFTGTSFSASYRQWSQRFSPKEHPVLALPGAPQFPVLE 180
QY 181 EHRPQKYMVNSDEMVKTEGAQIHAHLVRPVYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
DB 181 EHRPQKYMVNSDEMVKTEGAQIHAHLVRPVYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
QY 241 SPQCVGYSRSTAAPLMTMCLPDLKEIQRAVKLWVRSILDAQSVVYVATDSYVPELQQLF 300
DB 241 SPQCVGYSRSTAAPLMTMCLPDLKEIQRAVKLWVRSILDAQSVVYVATDSYVPELQQLF 300
QY 301 KGKVKVSLKPEVAQVDLYLIGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
DB 301 KGKVKVSLKPEVAQVDLYLIGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
QY 361 LRDEF 365
DB 361 LRDEF 365

RESULT 2
US-10-301-822-169
; Sequence 169, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHOD OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301.822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169
; LENGTH: 388
; TYPE: PRT
```

```

; ORGANISM: Homo Sapiens
US-10-301-822-169

Query Match          100.0%; Score 1979; DB 14; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.2e-195;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYCPCMGFRGNQADHFLGSLAFAPKLLNRTLAVPPWIEYQHKKPPTNL 60
DB 24 MPAGSWDPAGYLLYCPCMGFRGNQADHFLGSLAFAPKLLNRTLAVPPWIEYQHKKPPTNL 83
QY 61 HVSQYKFKLEPLQAYHRVVISLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKKTCPMK 120
DB 84 HVSQYKFKLEPLQAYHRVVISLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKKTCPMK 143
QY 121 EGNPFGPFWDQFHVSFNKSELFTGTSFSASYRQWSQRFSPKEHPVLALPGAPQFPVLE 180
DB 144 EGNPFGPFWDQFHVSFNKSELFTGTSFSASYRQWSQRFSPKEHPVLALPGAPQFPVLE 203
QY 181 EHRPQKYMVNSDEMVKTEGAQIHAHLVRPVYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
DB 204 EHRPQKYMVNSDEMVKTEGAQIHAHLVRPVYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 263
QY 241 SPQCVGYSRSTAAPLMTMCLPDLKEIQRAVKLWVRSILDAQSVVYVATDSYVPELQQLF 300
DB 264 SPQCVGYSRSTAAPLMTMCLPDLKEIQRAVKLWVRSILDAQSVVYVATDSYVPELQQLF 323
QY 301 KGKVKVSLKPEVAQVDLYLIGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
DB 324 KGKVKVSLKPEVAQVDLYLIGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 383
QY 361 LRDEF 365
DB 384 LRDEF 388

RESULT 3
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
```



```
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-18162

Query Match      42.0%; Score 831; DB 20; Length 402;
Best Local Similarity 44.7%; Pred. No. 9e-77; Indels 26; Gaps 9;
Matches 167; Conservative 62; Mismatches 119; Indels 26; Gaps 9;

Qy 7 DPAGYLPCPCMGRCQAADHFLGSLAFKLNRTLAVPPHIEYQHKKPPFTNLHVSQK 66
Db 28 DPGYLYTCCPCMGRCQAADHFLGSLAFKLNRTLAVPPHIEYQHKKPPFTNLHVSQK 66
Qy 7 YFLEPLQYHVRISLEDFMEKLAPTHWPKEKRVAYCFE-----VAORSPPDKTCTPKE 121
Db 67 YFLEPLQYHVRISLEDFMEKLAPTHWPKEKRVAYCFE-----VAORSPPDKTCTPKE 121
Qy 86 YFEVEPEKTHRVITNADFVHMLADDINFESEKRVFCYKERYSLQOEKNDPKPNCCHAD 145
Db 86 YFEVEPEKTHRVITNADFVHMLADDINFESEKRVFCYKERYSLQOEKNDPKPNCCHAD 145
Qy 122 GNFGFGFDORHVSFNKSELTFTGISASVREQ--WSORFSPKEHVPALPCAPAFPV 178
Db 146 GNFGFGFDORHVSFNKSELTFTGISASVREQ--WSORFSPKEHVPALPCAPAFPV 178
Qy 179 LEHRLPQKMYWSDMVKYTKTGAEQIHHLVR-PYGIHLRIGSDWKACAMLDGDTAGSH 237
Db 179 LEHRLPQKMYWSDMVKYTKTGAEQIHHLVR-PYGIHLRIGSDWKACAMLDGDTAGSH 237
Qy 206 QLECKLQYLOWSRYSREASKDFIREQLPRGAFGLHLENGIDWVRACEHVKD---SQH 262
Db 206 QLECKLQYLOWSRYSREASKDFIREQLPRGAFGLHLENGIDWVRACEHVKD---SQH 262
Qy 238 FMASPCQGVYSRSTAAPIWTMTMCLPDLKETQRAVKLVRSL-----DAQSVVYATDSE 290
Db 238 FMASPCQGVYSRSTAAPIWTMTMCLPDLKETQRAVKLVRSL-----DAQSVVYATDSE 290
Qy 263 LPASPCQGLYKNERGA-LYPELNPESKEALIRQLKTKNVRQTQPDNEIKSVFVADSN 321
Db 263 LPASPCQGLYKNERGA-LYPELNPESKEALIRQLKTKNVRQTQPDNEIKSVFVADSN 321
Qy 291 SYVPELO-QLFKGVKVVSLKPEVAQVDLYLGOADHFTGNCVSSTAFVKKRRDLQGRP 349
Db 291 SYVPELO-QLFKGVKVVSLKPEVAQVDLYLGOADHFTGNCVSSTAFVKKRRDLQGRP 349
Qy 322 HMIQELNTALSRWGISVHKLPEDDPYLDLALQSNHFNIGNCISSYSAFEKRRDVGFP 381
Db 322 HMIQELNTALSRWGISVHKLPEDDPYLDLALQSNHFNIGNCISSYSAFEKRRDVGFP 381
Qy 350 SSFEGMDRPPKLRD 363
Db 382 SYFWGP---PKED 392
```

```
RESULT 6
US-09-833-245-366 Application US/09833245
; Sequence 366, Publication No. US20040010341
; GENERAL INFORMATION:
; APPLICANT: Humantec Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: P5546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 366
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-366
```

```
Query Match      40.9%; Score 809; DB 11; Length 165;
Best Local Similarity 96.9%; Pred. No. 4.7e-75;
Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 206 HLVRPVGVIHLRIGSDWKACAMLDGDTAGSHFMASPCQGVYSRSTAAPIWTMTMCLPDLK 265
Db 6 HFARTYPIHLRIGSDWKACAMLDGDTAGSHFMASPCQGVYSRSTAAPIWTMTMCLPDLK 65
Qy 266 EIORAVKLWVRLSDAQSVVYATDSESVPELOQLFKGVKVVSLKPEVAQVDLYLGOAD 325
Db 66 EIORAVKLWVRLSDAQSVVYATDSESVPELOQLFKGVKVVSLKPEVAQVDLYLGOAD 125
Qy 326 HFTGNCVSSTAFVKKRRDLQGRPSSTFGMDRPPKLRDEF 365
Db 126 HFTGNCVSSTAFVKKRRDLQGRPSSTFGMDRPPKLRDEF 165

RESULT 7
US-09-774-954-8
; Sequence 8, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-Nov-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-Jan-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-774-954-8

Query Match      27.4%; Score 541.5; DB 11; Length 474;
Best Local Similarity 36.2%; Pred. No. 1.1e-46;
Matches 125; Conservative 60; Mismatches 107; Indels 53; Gaps 11;

Qy 19 GRFGNQADHFLGSLAFKLNRTLAVPPHIEYQHKKPPFTNLHVSQKYLEPQAVHR 78
Db 166 GRFGNQVDQLGLVLAFAKALDRLVLPNFIFFKH---PETQM-IPFEFLQVGTVAKTR 221
Qy 79 VISLEDFMEKLAPTHWPKEKRVAYCFEVAQRSDDKTCPMKEGNPFQGFWDQFVSNK 138
Db 222 VVTWQEFETKIMPTHFVGTFRQA-IYDKSABFG-----CHSKEGNPFQGFWDQIDVSVFG 275
```

139 SELFCTI--SFSASY---REWSQSPKEHPVIALPGAPOPVLEHRLQKYMVWSD 193
 276 DEYFGDIPGGFDLQNGSRKKLEKPESEYVPLAFSSAPPPSGKVKWSIQYLKRWSS 335
 194 EMVKTEAQIHAHLVRPYPYGIHLRIGSDWKNAACMLKDGDTAGSHFMSPPQCVGYSRSTAA 253
 336 RITEQAKKIPISANLAKFPFVAVHLRNDADWVRVCEHI-DITTNRPLFASQCLGEGHILGT 394
 254 PLMTWCLPDLKEIQRAVKLWVRSLDAQSVYVATDSYVPELQQLFKGKVKVUSLKPEV 313
 395 -LTKEICSPSKQI-----LEQ-----IEAHRQEPDD 420
 314 AQVDLVLGADHFGICNGVSSFTAFVKRERDLOG---RPSSRFGM 355
 421 MYTSLAIMGRADLFVGNVCVSTFTSHIVKRERDHAGQSPRPSAFPGI 465

RESULT 8

US-09-774-954-9
 ; Sequence 9, Application US/09774954
 ; Publication NO. US20040241645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/774,954
 ; FILING DATE: 30-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,741
 ; FILING DATE: 26-NOV-1997
 ; APPLICATION NUMBER: 08/792,498
 ; FILING DATE: 31-JAN-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1041P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 61 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
 ; US-09-774-954-9

Query Match 17.8%; Score 353; DB 11; Length 61;
 Best Local Similarity 100.0%; Pred. No. 1.8e-28;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MPAGSWDPAGVLLYCPGMRFGNQADHFLGSLAFKLNRTLVAVPPWIEYQHKKPPFTNL 60
 1 MPAGSWDPAGVLLYCPGMRFGNQADHFLGSLAFKLNRTLVAVPPWIEYQHKKPPFTNL 60
 61 H 61
 61 H 61

RESULT 9

US-09-774-954-3
 ; Sequence 3, Application US/09774954
 ; Publication NO. US20040241645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/774,954
 ; FILING DATE: 30-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,741
 ; FILING DATE: 26-NOV-1997
 ; APPLICATION NUMBER: 08/792,498
 ; FILING DATE: 31-JAN-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1041P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 61 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 ; US-09-774-954-3

Query Match 14.9%; Score 295; DB 11; Length 61;
 Best Local Similarity 91.5%; Pred. No. 1.8e-22;
 Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

3 AGSWDPAGVLLYCPGMRFGNQADHFLGSLAFKLNRTLVAVPPWIEYQHKKPPFTNLH 61
 3 AGSWDLAGVLLYXXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPPFTNLH 61

RESULT 10

US-10-471-450-26
 ; Sequence 26, Application US/10471450
 ; Publication NO. US20040152877A1
 ; GENERAL INFORMATION:
 ; APPLICANT: JACKSON, Jennifer L.; BAUGHN, Mariah R.;
 ; APPLICANT: SWARNAKAR, Anita; YUE, Henry;
 ; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
 ; APPLICANT: DING, Li; TANG, Y. Tom;
 ; APPLICANT: LEE, Soo Yeun; AZIMZAI, Yalda;
 ; APPLICANT: CHAWLA, Narinder K.; GIETZEN, Kimberly J.;
 ; APPLICANT: GRIFFIN, Jennifer A.; LAL, Preeti G.;
 ; APPLICANT: YANG, Junming; BOROWSKY, Mark L.;
 ; APPLICANT: RICHARDSON, Thomas W.; YUE, Huibin;
 ; APPLICANT: BECHA, Shanya; FORSYTHE, Ian J.;
 ; APPLICANT: JONES, Karen Anne; WARREN, Bridget;
 ; APPLICANT: THANGAVELU, Kavitha; HONCHELL, Cynthia;
 ; APPLICANT: JOLLEY, Helen E.; HAPALIA, April J.A.; Huijun Z. Ring

```
; TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0917 USN
; CURRENT APPLICATION NUMBER: US/10/471,450
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: PCT/US02/07869
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/276,857
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/285,489
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/285,556
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/288,700
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/288,646
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/290,510
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/290,369
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/332,426
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7503717CD1
US-10-471-450-26

Query Match      8.3%; Score 164.5; DB 16; Length 328;
Best Local Similarity 21.7%; Pred. No. 6.7e-08;
Matches 85; Conservative 41; Mismatches 115; Indels 151; Gaps 15;

QY 3 AGSWDPAG-----YLLY--CPGGRFGNQADHFLGSLAFAPAK 36
Db 13 AVSWPPASASGQFWPGQSAADILGSAARRRLLYDVNPPG-FNLRRDYIIIRIALSK 71
QY 37 LLNRT-----LAVPPMTEYOH-HKPPFTNLHVSQYKFKLEPLQAYHRVISLEDFMEKAP 91
Db 72 TLLKTEEWLVLPWPCRLYHQSPDIHQVRIPWSEFDFLPSLNKNIPVIEYEQPIAESG- 130
QY 92 THWPPKRVAYCFEVAQRSDDKTCPMKEGNPFQDFHVSFNKSGELFTGISFSASY 151
Db 131 -----GPFIDQVYV-----LQSY 143
QY 152 REQWSQRFSPKEHPVLALPCAPQFPVLEHRPQKXWV-----WSDMKVKTGEAQIH 206
Db 144 AEGWKE--GTWEEKV-----DERPCIDQLLYFQEDWKKVKVLGSA----- 182
QY 207 LVRPVYGIHLRIGSDWKNACAMLDKGTAGSHFMASPOCVGYSRSTAAPIWTMCLPOLKE 266
Db 183 LGGPYLGVHLR-----RKDFIWGHR-----QDVPSLEG 210
QY 267 IQRAVKLWVRSLDQAQSVYVATDSYVPELQQLFKGKVKVSLKPE-----VAQ 315
Db 211 AVRKIRSLMKTURLDKVVFATDAVR--KEYBELKLLPEWVRPEPTWBELELYKGGVAI 268
QY 316 VDLVILQADHFGNCVSSFTAFVKRRERDLOG 347
Db 269 IDQWICAHARFFIGTSVTSIFSFRIHEREILG 300

RESULT 11
US-11-097-143-36192
; Sequence 36192, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.

; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36192
; LENGTH: 490
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-36192

Query Match      8.1%; Score 159.5; DB 20; Length 490;
Best Local Similarity 22.4%; Pred. No. 4e-07;
Matches 92; Conservative 55; Mismatches 159; Indels 105; Gaps 19;

QY 9 AGYLLYPCPMGR-FGNQADHFLGSLAFAKLLNR-----TLAVPPMTEYOH-HKPPFTN 59
Db 71 AVYLLYDVNISEGNLRDVRVIRMAVFVRLQRRRRFRHVRVLVLPWPRLYHWHSQGLQQ 130
QY 60 LHSVYQYKFKLEPLQAYHRVISLEDFM--EKLAPTHWPPKRVAYCFEV----- 106
Db 131 SGLWSPHFFDLASRLRYAVFLVDYEEFLAEQRLFCNPGAPLVHVGHAFRLQHYEVMLEQGI 190
QY 107 ---AAQSPDKTCPMKEGN-PFGPFWDQFHVSFNKSELFTGISFSAS-----YREOW 155
Db 191 FRDKFERVTDK---PCEGSLGGPLLOQAEIRVGR---PHCVRFQGSAGLLEKLLREAI 244
QY 156 SQRFSPKEH-----PVIALPGAPQAPFPVLEHRPQKXWVSDMKVKTGEAQIHAIHLVPRY 211
Db 245 DEDTAGPEDVDDMRITVALLSAET---VLHDS-----WGDE-----HFWQARRS 284
QY 212 VGIHLR---IGSDWKNACAMLDKGT-----GSHFMASPOCVGVSR 249
Db 285 MEFARLEQVAADFRRQALDITDASAGVQRPAMWELERPKRNAKGGDYLCALHRLRGDFVR 344
QY 250 STAAPLWTMCLPDLKEIORAVKLVWRSLDQAQSVYVATDSY-VPELQQLFKGKVKVVS 308
Db 345 SDA-----TPTLKAAQOQVKQLLRGNMTTFLATDATPYELMELKELFY-RFRLVH 397
QY 309 LKPE-----VAQVDLYLQADHFGNCVSSFTAFVKRRERDLOG 347
Db 398 FAPESNVQRRELKDGGVAVVDQLVCAYARYFVGVESTFTYRIYEREILG 448

RESULT 12
US-10-471-450-8
; Sequence 8, Application US/10471450
; Publication No. US20040152877A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, Jennifer L.; BAUGHN, Mariah R.;
; APPLICANT: SWARNAKAR, Anita; YUE, Henry;
; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
; APPLICANT: DING, Li; TANG, Y. Tom;
; APPLICANT: LEE, Soo Yeun; AZINZAI, Yalda;
```

APPLICANT: CHAWLA, Narinder K.; GIETZEN, Kimberly J.;
 APPLICANT: GRIFFIN, Jennifer A.; LAL, Preeti G.;
 APPLICANT: YANG, Junming; BOROWSKY, Mark L.;
 APPLICANT: RICHARDSON, Thomas W.; YUE, HuiBin;
 APPLICANT: BECHA, Shanya; FORSYTHE, Ian J.;
 APPLICANT: JONES, Karen Anne; WARREN, Bridget;
 APPLICANT: THANGAVELU, Kavitha; HONCHELL, Cynthia;
 APPLICANT: JOLLEY, Helen E.; HAFALIA, April J.A.; Huijun Z. Ring
 TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
 FILE REFERENCE: PF-0917 USN
 CURRENT APPLICATION NUMBER: US/10/471,450
 CURRENT FILING DATE: 2003-09-10
 PRIOR APPLICATION NUMBER: PCT/US02/07869
 PRIOR FILING DATE: 2002-03-14
 PRIOR APPLICATION NUMBER: US 60/276,857
 PRIOR FILING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: US 60/285,489
 PRIOR FILING DATE: 2001-04-19
 PRIOR APPLICATION NUMBER: US 60/285,556
 PRIOR FILING DATE: 2001-04-19
 PRIOR APPLICATION NUMBER: US 60/288,700
 PRIOR FILING DATE: 2001-05-04
 PRIOR APPLICATION NUMBER: US 60/288,646
 PRIOR FILING DATE: 2001-05-04
 PRIOR APPLICATION NUMBER: US 60/290,510
 PRIOR FILING DATE: 2001-05-10
 PRIOR APPLICATION NUMBER: US 60/290,369
 PRIOR FILING DATE: 2001-05-11
 PRIOR APPLICATION NUMBER: US 60/332,426
 PRIOR FILING DATE: 2001-11-16
 NUMBER OF SEQ ID NOS: 52
 SOFTWARE: PERL Program
 SEQ ID NO 8
 LENGTH: 429
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: Incyte ID No: 7493913CD1
 US-10-471-450-8

Query Match 7.2%; Score 142; DB 16; Length 429;
 Best Local Similarity 22.08; Pred. No. 2.1e-05;
 Matches 99; Conservative 46; Mismatches 138; Indels 168; Gaps 19;
 QY 3 AGSWDPAG-----YLLY--CPCMGFRGNQADHFLGSLAFK 36
 DB 13 AVSWPPASAGQEPWPGQSAADILSGAASRRRLYLDVNPPEG-FNLRRDVIYRIASLLK 71
 QY 37 LLNRT-----LAVPPWIBYQH-HKPPFTNLHVSYQYKFKLEPQAYHRVLSLEDPMEKLP 91
 DB 72 TLLKTEEWLVLPWPWGLYHQSPDIHQVRIPWSEFFDLPSLNKNIPIVIEYEQF---IAE 128
 QY 92 THWPEKRVAYCEVAQRSPDKKTCPMKEGN-----PFGPFDQPHVSNKSELFTG-- 144
 DB 129 SGGPFIDQV-YVLQSYAE-----GWKETWEEKVDPERPCIDQLLYSQDKHYKRGWF 179
 QY 145 -----ISFSASY-----REQWSQRF 160
 DB 180 WGYEETRGVNSVCLSVQGSASIVAPLLRNTSARSVMDLRAENLLHDHGGKEWDFTRS 239
 QY 161 PKEHPVIALPGAPAQFPVLEHR-----PLQYMWWSDEWVKTGEAIIHAHL 207
 DB 240 -----MVFAHRLREVGDFRSRLNSTDDADRIPFQE--DWMKMKVGLGSA-----L 284
 QY 208 VRPVVGHILRGSDKNACAMLDGTAGSHFMAQPCVGVSRSTAAPLTMTMCLPDLKEI 267
 DB 285 GGPVILGVHLR-----RKDFIWHGR-----QDVPSLEGA 312
 QY 268 QRAVKLWRLSDAQSVVATDSESYVPELOQFKGKVKVWSLKPE-----VAQV 316
 DB 313 VRKIRSLMKTHRLDKVFVATDAVR--KEYBELKLLPEMVRFEPTWEELELYKDGGVAIL 370

QY 317 DLYILQADHFIGNCVSSFTAFVYKRRDLOG 347
 DB 371 DQWICAHARFFIGTSVSTSPRIHERELG 401
 RESULT 13
 US-10-820-474A-102
 Sequence 102, Application US/10820474A
 Publication No. US20050155089A1
 GENERAL INFORMATION:
 APPLICANT: LAL, PREETI
 APPLICANT: TANG, Y. TOM
 APPLICANT: GORGONE, GINA A.
 APPLICANT: CORLEY, NEIL C.
 APPLICANT: GUEGLER, KARL J.
 APPLICANT: BAUGHN, MARIAH R.
 APPLICANT: AKERBLOM, INGRID E.
 APPLICANT: YUE, HENRY
 APPLICANT: AU-YOUNG, JANICE
 APPLICANT: PATTERSON, CHANDRA
 APPLICANT: REDDY, ROOPA
 APPLICANT: HILLMAN, JENNIFER L.
 APPLICANT: BANDMAN, OLGA
 TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
 FILE REFERENCE: 039386-1568
 CURRENT APPLICATION NUMBER: US/10/820,474A
 CURRENT FILING DATE: 2004-04-07
 PRIOR APPLICATION NUMBER: 09/720,533
 PRIOR FILING DATE: 2001-03-20
 PRIOR APPLICATION NUMBER: PCT/US99/14484
 PRIOR FILING DATE: 1999-06-25
 PRIOR APPLICATION NUMBER: 60/090,762
 PRIOR FILING DATE: 1998-06-26
 PRIOR APPLICATION NUMBER: 60/094,983
 PRIOR FILING DATE: 1998-07-31
 PRIOR APPLICATION NUMBER: 60/102,686
 PRIOR FILING DATE: 1998-10-01
 NUMBER OF SEQ ID NOS: 269
 SOFTWARE: PatentIn version 3.3
 SEQ ID NO 102
 LENGTH: 150
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: Incyte Clone No: 1647884
 US-10-820-474A-102
 Query Match 6.1%; Score 120; DB 18; Length 150;
 Best Local Similarity 95.0%; Pred. No. 0.00087;
 Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPAGSWDPAGYLLYPCWGR 20
 DB 24 MPAGSWDPAGYLLYPCWGR 43
 RESULT 14
 US-10-424-599-145078
 Sequence 145078, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kovalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 145078

```

; LENGTH: 543
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_102023C.1.pep
US-10-424-599-145078

Query Match      5.4%; Score 117; DB 15; Length 543;
Best Local Similarity 22.0%; Pred. No. 0.012;
Matches 89; Conservative 50; Mismatches 137; Indels 128; Gaps 23;

QY 6 WPA-GYLLYCPMGRGNQADHFL---GSLAFKLNRTLAVP-PWIEYQHHK----- 54
DB 190 WNPCKGKFLAICVS---GQSNHLICLEKHIFPAALLNRVLVIPSCKVDYQYDRVVDIH 247
QY 55 -----PPPTNL---HVSQK---YFKLEPQOAHRVISLED-FMEKLPATHW 94
DB 248 INCLGKKVVVSFDFVSNLKKHGHLDKFLCYFS-QPQPCY-----LQDERLKKGALGL 301
QY 95 PPEKRVAYCFEVAQAQRSPDKTKCMKEGPNPFGFDQFHVSNKSELFTGISFSASYREQ 154
DB 302 TMSKPEAVNDE--DTRKPKKTVQ---DVLGKF-----SFDDVMAIGDVFYAEVERE 349
QY 155 WSQRFSPKEHPVLALPGAPAOF---PVLBEHRPLQKYMVMSDEMVKTGAEQIHAHLVREY 211
DB 350 WVMQ-----PGPIAHKCKTLIEPNRLI-----LTAQRFIQTFLGRNF 388
QY 212 VGIHLRIGSDWKNACAMLDGTAGSHFMASPOCVGYSRSTAAPLTMTCMLPDLKEIQRAV 271
DB 389 IALHFR-----RHGFLKFCNAKKFSCFYPIQAAAD-----CI--LRVVEMA- 427
QY 272 KLWVRSLSDAQSVYVATD--SESYPPELOQL--FKGKVKVSLKPE----- 312
DB 428 -----DAPILYSLTDAESETGLQSLVNLGRPVPLVIRPARNSAEKWDALLYRHNM 480
QY 313 -----VAQVDLYLQADHFIQNCVSSSTAFVKKRERDLQGRPS 350
DB 481 DGDSQVEALMDXTCAMSSVFAGPAGSTIFTDLRLRDKWGSAS 524

RESULT 15
US-10-437-963-162121
; Sequence 162121, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10437,963
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 162121
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61241C.1.pep
US-10-437-963-162121

Query Match      5.4%; Score 106.5; DB 16; Length 441;
Best Local Similarity 19.9%; Pred. No. 0.1;
Matches 77; Conservative 63; Mismatches 147; Indels 99; Gaps 17;

QY 4 GSWDPAGYLLYCPMGRF-----GNQADHFLGSLAFKLNRTLAVPWPWIEYQHH 53
```

```

DB 51 GFYDPTADLLPGIRRGKFEIVPQIIWGLNNQKIAPARACLTARFLNRSLLMPSLSASLFY 110
QY 54 KPFFTNLHVSQYKYPKLEPQOAHRVISLED-FMEKLPATHWPPPEKRVAYCFEVAQAQRSPD 113
DB 111 KEYVDLLRPITFDKVFDFTKFNA-----RCQGFV-----RLARYSEVSNQTKPF 153
QY 114 KKTCPMKEGPNPFGFDQFHVSNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGA- 172
DB 154 K-----LOKGS--GREWT---VEKDLDQLL-----QVR-----RGEADDSEVIEIIGH 192
QY 173 PAQFPVLEBHRPLQKYMVMSD-----EMVKT---GEAQIHA-----HLV 208
DB 193 PDLWP---DHWEVKDYARIFDCLALVPEIETEVVKAISKIREAGIKARHEAGISHNKHVK 249
QY 209 R-----PYVGIHLRIGSDWKNACAMLDGTAGSHFMASPOCVGYSRSTAAPLTMTCML 261
DB 250 DGTMPNPPFYIAVHRIEKDMWIHCKWQORSNKEICSSKEEIIHKVS----- 298
QY 262 PDLKEIQRAVKLWVRSlda--OSVYVATDSSESYPPELOQLFKGKVKVSLKPEV--AQVD 317
DB 299 -QITDLRRPVVVLAVADSLLEDSDITSGRVGMVAPEKKRLGVTDIYNROPYLIKSAID 357
QY 318 LYILQADHFIQNCVSSSTAFVKKRER 343
DB 358 FEVCARADVFVGNSEFSTFNSNLVLGR 383
```

Search completed: October 25, 2005, 15:41:21
Job time : 216.353 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 03:15:37 ; Search time 1022.3 Seconds
(without alignments)
2113.573 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLCPMGR.....QGRPSFFGMDRPPKLRDF 365

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool.h/US09774954/runat_25102005_105431_6380/app_query.fasta_1.917
-DB=N-Geneseq -QFMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.1 703 @runat_25102005_105431_6380 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	100.0	1514	2	AAV65632 Human hea
2	1979	100.0	5218	13	Acn40746 Tumour-as
3	1979	100.0	5266	10	Adf81754 Leukaemia
4	1979	100.0	11284	2	AAV65633 Plasmid c
5	1867.5	94.4	5230	4	Aak51510 Human pol

6	1855.5	93.8	4850	4	AAK52494	Aak52494 Human pol
7	1844	93.2	5009	2	AAV65634	AAV65634 First Eco
8	831	42.0	1209	4	ABL13947	ABL13947 Drosophil
9	811	41.0	610	13	ADQ56824	Adq56824 Novel can
10	809	40.9	515	4	AAU07629	Aad07629 Human sec
c 11	796	40.2	3264	4	ABL13946	ABL13946 Drosophil
c 12	506.5	25.6	3793	4	ABL13894	ABL13894 Drosophil
13	424	21.4	10331	4	AAK79915	Aak79915 Human imm
14	424	21.4	10331	4	AAU03344	Aal03344 Human rep
15	393	19.9	479	9	ACH32225	Ach32225 Human end
c 16	382	19.8	3567	4	ABL13952	ABL13952 Drosophil
17	353	17.8	477	9	ACH35532	Ach35532 Human end
18	219	11.1	1545	4	AAF23894	Aaf23894 Human sec
19	169.5	8.6	2229	9	ADA21149	Ada21149 Human sec
20	166.5	8.4	2151	9	ADA21151	Ada21151 Human sec
21	164.5	8.3	1967	6	ABS73886	Abst73886 Human CDN
22	164	8.3	1473	13	ADQ97980	Adq97980 Human can
23	164	8.3	1926	4	ABL25967	ABL25967 Drosophil
c 24	164	8.3	3926	4	ABL25966	ABL25966 Drosophil
25	149.5	7.6	2374	9	ADA21150	Ada21150 Human sec
26	142	7.2	2434	6	ABS73868	Abst73868 Human CDN
27	120	6.1	760	3	AZ98210	Aaz98210 Human sig
c 28	115.5	5.8	4104	4	ABL25964	ABL25964 Drosophil
29	106	5.4	2365	9	ADA21178	Ada21178 Human sec
30	105.5	5.3	4845	3	AACT4541	Aac74541 Human ORF
31	104	5.3	2071	3	AACT5262	Aac52262 Arabidops
32	104	5.3	97081	12	ADQ97980	Adq97980 Human can
33	103.5	5.2	960	10	ADQ95589	Adq95589 Human NOV
34	103.5	5.2	960	10	ADQ95587	Adq95587 Human NOV
35	102.5	5.2	783	8	ABT19142	Abt19142 Aspergill
36	102.5	5.2	907	8	ABT18548	Abt18548 Aspergill
37	102.5	5.2	2162	4	AAK73291	Aak73291 Human imm
38	102.5	5.2	2907	8	ABT17954	Abt17954 Aspergill
39	102	5.2	2069	3	AACT5455	Aac54555 Arabidops
40	102	5.2	29069	3	AAAB1497	Aaa81497 N. mening
41	102	5.2	110000	3	AAAB1490_11	Continuation (12 o
42	102	5.2	349980	3	AAF21609	Aaf21609 Neisseria
43	101.5	5.1	2366	1	AAQ91052	Aan91052 Phage DNA
44	101.5	5.1	2381	2	AAQ05241	Aaq05241 Sequence
45	101.5	5.1	2381	10	ADB57903	Adb57903 Toxicity-

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.

XX
AC
AAV65632;

XX
DT 16-DEC-1998 (first entry)

XX
XX Human heart O-fucosyltransferase encoding DNA.

XX
DE O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.

XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT misc_feature 1..1100

FT /*tag= b

FT /note= "this actively expressed O-fucosyltransferase
sequence is claimed for in claim 9"

FT CDS 1..1098

FT /*tag= a

FT /product= "human heart O-fucosyltransferase"

XX
PN WO9833924-A1.

XX
PD 06-AUG-1998.

XX
XX AAV65633 Plasmid c

PF 17-DEC-1997; 97WO-US023401.

XX 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
XX Wang Y, Spellman MW;
XX MPI: 1998-437477/37.
DR P-PSDB; AAM80571.
XX Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX Claim 9; Fig 12A; 90pp; English.
XX This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EGF domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC Probes based on EGF domain polypeptide are used to detect gene
CC amplification and expression. The expression can also be determined at
CC the protein level using antibodies specific for O-fucosyltransferase
XX SQ Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3, 08e-200 Length: 1514
Score: 1879.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-2 (1-365) x AAV65632 (1-1514)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrIleuLeuTyrCysProCysMetGlyArg 20
DB 1 ATGCCCGCGGGCTCTGGGACCCGGCGGTACTCTCTACTGCTCCCTCATGCGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerIleuAlaPheAlaLysLeuLeuAsnArg 40
DB 61 TTTGGGAACCCAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProTrpIleGlyTyrGlnHisIleLysProPheThrAsnLeu 60
DB 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACCAGCATCACAGCCTCTCTTACCACCTC 180
QY 61 HisValSerTyrGlnIleTyrPheLysLeuGluProLeuGlnAlaTyrHisArgVal 80
DB 181 CATGTCTCTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 240
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTyrProGluLysArgVal 100
DB 241 AGCTTGGAGGATTTCAAGGAGAGCTGGCACCACCTGCCCCCTTGGAGAGCGGGTG 300
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysIleThrCysProMetLys 120
DB 301 GCATCTACTGCTTGGTGGCCAGCCAGGAGCCAGTAAGAGACGTGCCCTCATGAG 360
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
DB 361 GAAGGAAACCCCTTTGGGCCCATTTCTGGGATCAGTTTCTCATGTGAGTTTCAACAGTCGGAG 420
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
DB 421 CTTTTTACAGGCATTTCTTCACTGCTTCTTACAGAGAACATGAGAGCCAGAGATTTTCT 480
QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180

DB 481 CCAAAGGAACATCCCGTCTTGGCCCTGCCAGGAGCCCGCCAGTTCCTCCCGTCTTAGAA 540
QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
DB 541 GAACACAGCCCACTACAGAGTACATGTATGTCTAGACGAAATGTGTGAAGACGGAGAG 600
QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
DB 601 GCCCAGATTATCCGCCCACTTGTCCGGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660
QY 221 AspTyrIleAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
DB 661 GACTGGAAGACCGCTGTGCCATCTGAAGGACGGGATGCGAGGCTCGCATTCTATGGCC 720
QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
DB 721 TCTCCGAGTGTGTGGGCTACAGCCGCGAGCACAGCGGCCGCCCTCACGATGACTATGTGC 780
QY 261 LeuProAspLeuLysGluIleGlnArgAlaValIleValLeuTyrValArgSerLeuAspAla 280
DB 781 CTGGCTGACTGAGAGAGANTCAGAGGGCTGTGAAGCTCTGGGTGAGTCTGCTGATGCC 840
QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
DB 841 CAGTCGGTCTACCTTGTCTACTGATTCGAGATTCGAGATTAATGTGCTGAGCTCCACACCTTC 900
QY 301 LysGlyLysValIleValIleValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
DB 901 AAAGGGAGGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCGAGTGCACCTGACTGATC 960
QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
DB 961 CTCGGCCAGCCGAGCACCTTTATTGGCAACTGTGTCTCTCTCTTCTTCTTCTTCTTCTTCT 1020
QY 341 ArgGluArgAspLeuGlnGlyValArgProSerSerPhePheGlyMetAspArgProLys 360
DB 1021 CGGAGCGGGACCTCCAGGGAGGCGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1080
QY 361 LeuArgAspGluPhe 365
DB 1081 CTGCGGAGCAGGTTTC 1095

RESULT 2
ACN40746
ID ACN40746 standard; cDNA; 5218 BP.
XX ACN40746;
AC ACN40746;
XX 18-NOV-2004 (first entry)
DE Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX WO2004030615-A2.
XX 15-APR-2004.
XX 29-SEP-2003; 2003WO-US028547.
XX 02-OCT-2002; 2002US-0414971P.
XX (GETH) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
PI

XX WPI; 2004-347921/32.
 DR P-PSDB; ABM82223.
 XX
 XX New tumor-associated antigenic target polypeptides and nucleic acids,
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 XX
 XX Claim 1; SEQ ID NO 5725; 7273pp; English.
 PS
 XX The invention relates to human tumour-associated antigenic target (TAT)
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are
 CC overexpressed in cancer tissues compared to normal tissues, and may thus
 CC serve as effective targets for the diagnosis and treatment of cancer in
 CC mammals. The invention also relates to the TAT nucleic acids and
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT nucleic acid of the invention
 XX
 XX Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 1,89e-199 Length: 5218
 Score: 1979.00 Matches: 365
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 13 Gaps: 0

US-09-774-954-2 (1-365) x ACN40746 (1-5218)

Qy 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 Db 119 ATGCTCGGGCTCCTGGAGCCCGCGGTACCTGCTCTACTGCTCCCTGCATGGGGCGC 178
 Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArg 40
 Db 179 TTTGGGAACACAGGCGCGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 238
 Qy 41 ThrLeuAlaValProTrpTrpIleGluTyrGlnHisLysProPheTrpAsnLeu 60
 Db 239 ACCTTGCTGCTCCTCTGGATTGAGTACCAGCATCACAGGCTCTTCCACCAACCTC 298
 Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValle 80
 Db 299 CATGTGCTTACCAGAAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 358
 Qy 81 SerLeuGluAspPheMetGluLysLeuAlaPheTrpHisTrpProGluLysArgVal 100
 Db 359 AGCTTGGAGGATTTTCATGGAGAGCTGGACCCACCCACCTGGCCCTTGAGAAGCGGGTG 418
 Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 Db 419 GCATCTGCTTTTGGTGGAGCCGAGGAGCCAGCAAGAGAGAGCTGCCCATGAG 478
 Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
 Db 479 GAAGGAACCCCTTTGGGCCATTTCTGGATCAGTTTCTATGTGAGTTTCAACAAGTCGGAG 538
 Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160

Db 539 CTTTTCACAGGCAATTCCTTCAGTGTCTTCTCAGAGAACAAATGGAGCCAGAGATTTCT 598
 Qy 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
 Db 599 CCAAGGAACATCCGGTGTCTGCCCTGGCAGAGAGCCAGCCAGGTTCCCGTCTTAGAG 658
 Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 Db 659 GAACACAGGCGCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 718
 Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 Db 719 GCCAGATTCATGCCCACTTGTTCGGCCCTATGTGGCATTCATTCGCGCATTTGGCTCT 778
 Qy 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 Db 779 GACTGGAAAGCGCTGTGCCATGCTGAAGAGCGGACTGCGAGGCTCGCAGCTTCATGGCC 838
 Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
 Db 839 TCTCCGAGTGTGTGGCTACAGCCGACAGCGGCGCCCTCACGATGACTATGTGC 898
 Qy 261 LeuProAspLeuLysGluLeuGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
 Db 899 CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 958
 Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
 Db 959 CAGTCGGTCTACGTTCTACTGATTCGAGAGTTATGTGGCTGAGCTCCAAAGCTCTTC 1018
 Qy 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
 Db 1019 AAAGGAAGGTGAAGTGTGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGATCATC 1078
 Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
 Db 1079 CTGGCCAAAGCCGACCACTTATTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1138
 Qy 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
 Db 1139 CGGGAGCGGACCTCCAGGGGAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1198
 Qy 361 LeuArgAspGluPhe 365
 Db 1199 CTGGGAGCAGATTC 1213

RESULT 3
 ID ADF81754 standard; DNA; 5266 BP.
 XX ADF81754;
 XX
 XX 26-FEB-2004 (first entry)
 XX
 XX Leukaemia-related DNA sequence #2310.
 XX
 XX Cytostatic; Gene therapy; leukaemia; ss.
 XX
 XX Unidentified.
 XX
 XX WO2003039443-A2.
 XX
 XX 15-MAY-2003.
 XX
 XX 04-NOV-2002; 2002WO-EP012303.
 XX
 XX 05-NOV-2001; 2001EP-00126244.
 XX
 XX 30-APR-2002; 2002EP-00009758.
 XX
 XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 XX (UFLU-) UNIV LUDWIG MAXIMILIANS.
 XX (HAFE/) HAERLACH T.

201	AlaGlnIleHisAlaHisIleuValArgProTyrValGlyIleHisIleuArgIleGlySer	220
732	GGCCAGATTATCATCCCACTCTGTGCCGCCCTATATGGGCATTCATCTGGCGATTGGCTCT	791
221	AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
792	GACTCGAAGAACCCCTGTGTCATCTGAAGGACGGGACTGCAGGCTCGCACTTCATATGGCC	851
241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
852	TCTCCGCAAGTGTGTGGGCTCAGACCGAGCACAGCGGCCCCCTCAGCATGACTATGTGC	911
261	LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla	280
912	CTGCTCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGGTATGCC	971
281	GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe	300
972	CAGTCGGTCTACGTTGCTACTGATTCGAGAGTATATGTGCTCTGAGCTCCCAACAGCTCTC	1031
301	LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle	320
1032	AAAGAGAAAGTGAAGTGGTGTGAGCTCGAAGCTTGAGTGGCCAGGTCGACCTGTATCATC	1091
321	LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys	340
1092	CTCGGNCAAGCGGACCACTTTATGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTGAG	1151
341	ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys	360
1152	CGGGAGCGGACCTCCAGAGGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1211
361	LeuArgAspGluPhe	365
1212	CTCGGGGACGAGTTTC	1226
RESULT 4		
DAVE5633		
D	AAV5633	standard; DNA; 11284 BP.
CX	AAV5633;	
CX	AAV5633;	
CX	16-DEC-1998	(first entry)
CX	Plasmid construct for expression of human O-fucosyltransferase.	
CX	O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;	
CX	O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.	
CX	Homo sapiens.	
CX	Synthetic.	
Key	Location/Qualifiers	
FFH	4140..5333	
CDS	/*tag= a	
FFH	/*note= "insert coding for human O-fucosyltransferase."	
FFH	4218..4235	
FFH	/*tag= b	
FFH	/*note= "polyhistidine tag"	
FFH		
XX	W09833924-A1.	
XX		
XX	06-AUG-1998.	
XX		
XX	17-DEC-1997;	
XX	97WO-US023401.	
XX		
XX	31-JAN-1997;	
XX	97US-00792498.	
XX	26-NOV-1997;	
XX	97US-00978741.	
XX		
XX	(GETH) GENENTECH INC.	
XX		
XX	Wang Y, Spellman MW;	
XX		

XX WPI: 1998-437477/37.
 DR P-PSDB; AAW80573.
 XX
 PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.
 XX
 PS Example; Page 49-57; 90pp; English.
 XX
 CC This represents the nucleotide sequence of the plasmid construct used for
 CC the expression of human heart O-fucosyltransferase. The human O-
 CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
 CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
 CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
 CC domains, are used in diagnosis and treatment of conditions associated
 CC with overexpression of O-fucosyltransferase, to promote survival of
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
 CC used to detect gene amplification and expression. The expression can also
 CC be determined at the protein level using antibodies specific for O-
 CC fucosyltransferase
 XX
 SQ Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;
 XX

Alignment Scores:
 Pred. No.: 5,85e-199 Length: 11284
 Score: 1979.00 Matches: 365
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-09-774-954-2 (1-365) x AAV65633 (1-11284)

QY	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	4236	ATGCCCGGGCTCTGGAGCCCGCGGTACTGCTCTACTGCTCTGCTCTGATGGGGCGC	4295
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	4296	TTTGGGAACAGCGCGATCATCTTCTGGGCTCTGCGCATTTGCAAGCTGTAAACCGT	4355
QY	41	ThrLeuAlaValProProTyrPheGluTyrGlnHisLysProPheThrAsnLeu	60
DB	4356	ACCTTGGCTGCTCCCTCTGGATGATGATACAGATCAGCATCAGGCTCTTTCACCAACCTC	4415
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle	80
DB	4416	CATGTGTCTTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC	4475
QY	81	SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal	100
DB	4476	AGCTTGGAGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAAGCGGGTG	4535
QY	101	AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys	120
DB	4536	GCAATACTGCTTTGAGTGGGAGCCAGCAGCCAGATAGAGAGCTGCCCATGAAG	4595
QY	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140
DB	4596	GAAGAAACCCCTTTGGCCCATCTCGGATCAGTTTCAATGAGTTTCAACAGTCGGAG	4655
QY	141	LeuPheThrGlyLysSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer	160
DB	4656	CTTTTACAGGCATTTCTTCAGTGTCTTCTACAGAGAACAATGGAGCCAGAGATTTTCT	4715
QY	161	ProLysGluHisProValLeuAlaLeuProGlyValaProAlaGlnPheProValLeuGlu	180
DB	4716	CCAAAGGAACATCCGGTGTCTGCGAGAGCCCGCCAGCTTCCCGCTCTAGAG	4775
QY	181	GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu	200
DB	4776	GAACACAGGCCACTACAGAGTACATGATGTTGTCAGACGGAATGGTGAAGCGGGAGAG	4835

QY	201	AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer	220
DB	4836	GCCAGATTTCATGCCACCTTGTCCGCCCTATGTGGCATTTCATCTGCGCATTTGCTCT	4895
QY	221	AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
DB	4896	GACTGGAGAAGCCCTGTGCCATGTCTGAAGAGCGGACTGCAGGCTCGCATTTCATGGCC	4955
QY	241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
DB	4956	TCTCCGACATGTGTGGCTTACAGCCCGCAGACAGCGCCCCCTCAGATGATATGTGC	5015
QY	261	LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla	280
DB	5016	CTGCGCTGACTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCTGGATGCC	5075
QY	281	GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe	300
DB	5076	CAGTCGGTCTAGCTTCTACTGATTCGAGAGTTATGTGCTGAGCTCCAAACAGCTCTTC	5135
QY	301	LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle	320
DB	5136	AAAGGAAGGTGAAGTGTGAGCTGAAGCTTGAGTGGCCCGGCTGACCTGTATCATC	5195
QY	321	LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys	340
DB	5196	CTCGGCAAGCCGACCATTTATTGGCAACTGTGTCTCTCTCTTCACTGCTTTGTGAAG	5255
QY	341	ArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArgProProLys	360
DB	5256	CGGAGCGGACCTCCAGGGAGCGGTCTTCTTCTTCTGGCATGGACAGGCCCTTAAG	5315
QY	361	LeuArgAspGluPhe	365
DB	5316	CTGCGGACGAGTTC	5330
RESULT 5			
ID	AAKS1510	standard; cDNA; 5230 BP.	
XX	AAKS1510;		
XX	06-NOV-2001	(first entry)	
XX	Human polynucleotide SEQ ID NO 55.		
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;		
KW	nervous system disorder; arthritis; inflammation; ss.		
XX	Homo sapiens.		
XX	WO200157190-A2.		
XX	09-AUG-2001.		
XX	05-FEB-2001; 2001WO-US004098.		
XX	03-FEB-2000; 2000US-00496914.		
PR	27-APR-2000; 2000US-00560875.		
PR	20-JUN-2000; 2000US-00598075.		
PR	19-JUL-2000; 2000US-00620325.		
PR	01-SEP-2000; 2000US-00654936.		
PR	15-SEP-2000; 2000US-00663561.		
PR	20-OCT-2000; 2000US-00693325.		
PR	30-NOV-2000; 2000US-00728422.		
XX	(HYSE-) HYSEQ INC.		
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;		
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR P-PSDB; AAW78377.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 XX Claim 1; Page 638-642; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3656 (AAW80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.4e-187 Length: 5230
 Score: 1867.50 Matches: 350
 Percent Similarity: 96.98% Conservative: 3
 Best Local Similarity: 96.15% Mismatches: 6
 Query Match: 94.37% Indels: 5
 DB: 4 Gaps: 1

US-09-774-954-2 (1-365) x AAK51510 (1-5230)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet-----GlyArgPhe 21
 DB 160 GACATGGCCACGTGGCTACACATCTCAAGCTGTGGAGAGAGGAGCGGCGCTTT 219
 QY 22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeuLeuAsnArgThr 41
 DB 220 GGGAAACCAAGCCGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACC 279
 QY 42 LeuAlaValProProTrpIleGluTyrGlnHisIleValProPheThrAsnLeuHis 61
 DB 280 TTGGCTGTCTCTTGGATTGAGTACCAAGCATCAGAGCTCTTCCACCAACCTCCAT 339
 QY 62 ValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSer 81
 DB 340 GTGTCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGTCATCAGC 399
 QY 82 LeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgValAla 101
 DB 400 TTGGAGGATTTTCATGGAGAGCTGGCACCACCCACTGGCCCCCTGAGAGCGGTGACA 459
 QY 102 TyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
 DB 460 TACTGCTTTGAGTGGCAGCCAGCCAGCCAGATGAAGAGACGTGCCCTCATGAAGGAA 519
 QY 122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
 DB 520 GGAACACCTTTGSCCTCTTGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTT 579
 QY 142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerPro 161
 DB 580 TTTTACGCAATTCCTTCAGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCTCCA 639
 QY 162 LysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGlu 181
 DB 640 AAGGAACATCCGGTGTCTGCTGCTGCGAGAGCCCGCCAGCTTCCCGTCTCTAGAGAA 699
 QY 182 HisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAla 201

DB 700 CACAGGCCACTACAGAGTACATGGTATGGTACAGACGAATGGTCAAGACGGGAGAGGCC 759
 QY 202 GlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAsp 221
 DB 760 CAGATTTCATGGCCACTTGTTCGGGCCCTATGTGGGCATTCATCTGGCATTTGGCTCTGAC 819
 QY 222 TrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSer 241
 DB 820 TGAAGAACCGCTGTGCCATCTCTCAAGAGACGGGACTGCGAGCTCGCATTTTCATGGCTCT 879
 QY 242 ProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeu 261
 DB 880 CCGCATGTGTGGCTACAGCCGCGCAGCACAGCGGCCCTCCACGATGACTATGTGCTG 939
 QY 262 ProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGln 281
 DB 940 CTTGACCTGAAGAGATCCAGAGGGCTGTGAGCTCTGGGTGAGGTGGATGCCAG 999
 QY 282 SerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPheLys 301
 DB 1000 TCGGTCTACGTGTGCTACTGATTCGAGGTTATGTGCTGAGCTCCACAGCTCTTCAAA 1059
 QY 302 GlyLysValLysValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeu 321
 DB 1060 GGGAGGTGAAGTGTGTGAGCTGAAGCTGAGGTGGCCAGTGCCTGTGATCTCTCATCTC 1119
 QY 322 GlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArg 341
 DB 1120 GGCCAGCCGACCATCTTATTGGCAACTGTGTCTCTCTCTTCACTGCTTTGTGAAGCGG 1179
 QY 342 GluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLysLeu 361
 DB 1180 GAGCGGACCTCCAGGGAGGCGCTCTCTTTCTTGGCATGAGAGGCCCCCTAAAGCTG 1239

RESULT 6
 AAK52494
 ID AAK52494 standard; cDNA; 4850 BP.
 XX AAK52494;
 AC AAK52494;
 DT 06-NOV-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 2023.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX Homo sapiens.
 OS
 XX W0200157190-A2.
 PR
 XX 09-AUG-2001.
 PD
 XX 05-FEB-2001; 2001WO-US004098.
 PF
 XX 03-FEB-2000; 2000US-00496914.
 PR
 XX 27-APR-2000; 2000US-00560875.
 PR
 XX 20-JUN-2000; 2000US-00598075.
 PR
 XX 19-JUL-2000; 2000US-00620325.
 PR
 XX 01-SEP-2000; 2000US-00654936.
 PR
 XX 15-SEP-2000; 2000US-00663561.
 PR
 XX 20-OCT-2000; 2000US-00693325.
 PR
 XX 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 PA

PT domains - useful for diagnosis and treatment of diseases involving
PI overexpression of the enzyme.

XX Example 1; Fig 11; 90pp; English.

PS This represents a first EcoRI nucleotide fragment of human KIA0180. This
CC 5009 basepairs partial cDNA encodes for a protein of unknown function
CC from myeloblast cell line KG-1. The invention provides a human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase

SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4,18e-185 Length: 5009
Score: 1844.00 Matches: 343
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 93.18% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-2 (1-365) x AAV65634 (1-5009)

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
DB 2 AACGAGGCGGATCACTTCTGGGCTCTCTGGCAATTCGAAAGCTGCTAAACCGTACCTTG 61
QY 43 AlaValProTrpIleGluTyrGlnHisHisLeuProPheThrAsnLeuHisVal 62
DB 62 GCTGTCCTCCTTGGATTGAGTACAGCATCACAGCCCTCTTCCACCACTCCATGTG 121
QY 63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
DB 122 TCCTACCAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATCAGCTTG 181
QY 83 GluAspPheMetGluLysLeuAlaProThrHisTyrProGluLysArgValAlaTyr 102
DB 182 GAGGATTTTCATGGAGNAGCTGGCCACCCACCATCTGGCCCCCTGAGNAGCGGTGGCATAC 241
QY 103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
DB 242 TGCTTTGAGTGCGAGCCAGCCAGAGCCAGATAGAGACGCTGCCCATGAAGGAAGGA 301
QY 123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
DB 302 AACCCCTTTGGCCCATCTCGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 361
QY 143 ThrGlyLysSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
DB 362 ACAGGCATTTCTCTAGTGCTTCTACAGAGAACATGGAGCCAGAGATTTCTCCAAAG 421
QY 163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluHis 192
DB 422 GAACATCCGGTGCTTGCCCTGCGAGAGCCCGAGCCAGTCTCCGCTCTAGAGAAACAC 481
QY 183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
DB 482 AGGCCATACAGAAGTACATGTTATGTTGTCAGACCAATGTTGAAAGCAGCGGAGGCCAG 541
QY 203 IleHisAlaHisLeuValArgProTyrValGlyLysLeuArgGlyLeuAspTyr 222
DB 542 ATTATGCCCCACCTTGTCCGCCCTTATGTGGGCATTCATCTGCGCATTTGGCTCTGACTGG 601
QY 223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetLeuPro 242
DB 602 AAGAACCCCTGTGCCATGCTGAAAGACGGGACTGACGGCTCGCACTTCTATGGCTCTCCG 661

QY 243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
DB 662 CAGTGTGTGGGCTTACAGCCGCGACACAGCGGCCCTCCAGATGACTATGTGCTGCT 721
QY 263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
DB 722 GACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCCGTGGATCCCACTG 781
QY 283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPheLysGly 302
DB 782 GTTACGCTTGCTACTGATTCGAGAGTTTATGTGCTGAGCTCCAAACAGCTCTTCAAAGG 841
QY 303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
DB 842 AAGGTGAAGGTGTGTGAGCTGGAAGCTGAGGTGGCCCATGACCTGTACATCTCGGC 901
QY 323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
DB 902 CAGCGGACCACTTTATGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 961
QY 343 ArgAspLeuGlnArgProSerSerPhePheGlyMetAspArgProPolysLeuArg 362
DB 962 CGGGACCTCCAGGGAGGCGCTCTTCTTCTTCGGCATGGACAGGGCCCCCTAAGCTGCG 1021
QY 363 AspGluPhe 365
DB 1022 GACGAGTTC 1030
RESULT 8
ABLI3947
ID ABLI3947 standard; cDNA; 1209 BP.
XX ABLI3947;
AC ABLI3947;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmacological; gene; ss.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR P-PSDB; AB569844.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI1840-ABLI16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at fcp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;
 Alignment Scores: 3 58e-78 Length: 1209
 Pred. No.: 831.00 Matches: 167
 Score: 61.23% Conservatives: 62
 Best Local Similarity: 44.65% Mismatches: 119
 Query Match: 41.99% Indels: 26
 DB: 4 Gaps: 9
 US-09-774-954-2 (1-365) x ABL13947 (1-1209)
 QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
 DB 82 GATCCCAATGGCTACCTACCTACTGCTGCTGATGGAGCGCTTGGCAACACGAGCGGAC 141
 QY 27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
 DB 142 CACTTCTGGGATCATTTGGCTTTCGCAAGCGCTTAATCGACCCCTGATCTCGCGCGG 201
 QY 47 TrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHisValSerTyrGlnLys 66
 DB 202 TGGGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255
 QY 67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMet 86
 DB 256 TATTTGAAGTGGAGCCCTGAAGGATACCATCGCTCATCCATGGCAGATTTTCATG 315
 QY 87 GluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu--- 105
 DB 316 TGGCACCTGGCCGACGACATTTGGCCAGAAATCGAGCGAGTGTCTTTTCTCAAGAA 375
 QY 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
 DB 376 CGATATAGCTTCAGAGGAGAGAGAGATCCAGCAAGCGCAATGCGCAGCGCAGGAT 435
 QY 122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
 DB 436 GGCAATCTTTTGGTCCCTTTGGGACACATTTTCCATAGATCTTTGGCGGTGAGAGTTC 495
 QY 142 PheThrGlyLysSerPheSerAlaSerTyrArgGluGln-----TrpSerGlnArg 158
 DB 496 TATGCGCCACTTCATTTTGTATGTGTCATATAGCAACGAGGCTGCAAGTGGCAGACAAA 555
 QY 159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
 DB 556 TATCCTCGAAGATCATATATCCGCTACTCGGTTCACCGAGCTCCGGCTAGTTTTCCTGTT 615
 QY 179 LeuGluGluHisArgProLeuGlnLysTrpMetValTrpSerAspGluMetValLysThr 198
 DB 616 CAGCTAGAGAACTGCAAGCTGCAGCGTACTTGCAGTGGAGTCAACGATATAGGGAAGCA 675
 QY 199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
 DB 676 TCTAAGGATTTTCCCGAGAGAGCTGCTGCGGGTGCCTTTTGGGCAATTCATCTCGC 735
 QY 218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237
 DB 736 AACGGTATCGATTGGGTGAGAGCTGTGAGCAGCCTCAAGGAT-----AGCCAGCAT 786
 QY 238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
 DB 787 CTGTTTGGCTCGCGCAGTGTCTGGCTATATAAATGACGTGGTGA---CTTACCCG 843
 QY 258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
 DB 844 GAGCTCTGCATGCCCTCCAGGAGGCGCATCATCGCCAGCTTAAAGAGAACCATTAAGAAC 903
 QY 278 Leu-----AspAlaGlnSerValTyrValAlaThrAspSerGlu 290

DB 904 GTGCGGCAAACTCAGCGGACAAACAAATCAGTTTTCGTGGCGTCAGACTCCAAT 963
 QY 291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValValSerLeu 309
 DB 964 CACATGATTGGTGAACATAACACGCGCCCTTAGTCGCATGGGCACTAGTGTGCACAGCTG 1023
 QY 310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
 DB 1024 CCGGAGATGATCCTTACCTGGACTTGGCACTTTCGGACAGTCCGAACCACTTTATCGGC 1083
 QY 330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
 DB 1084 AACTGTATATCTCTTACTCGGCATTCGAAAGGAGGACGAGATGTGCACGGTTCCTCA 1143
 QY 350 SerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
 DB 1144 TCGTACTTCTGGGATTC-----CCACAGAAAGGAT 1176
 RESULT 9
 ADQ56824
 ID ADQ56824 standard; DNA; 610 BP.
 XX ADQ56824;
 XX 21-OCT-2004 (first entry)
 XX Novel canine microarray-related DNA sequence SeqID8126.
 XX canine microarray; drug screening; toxicity assay;
 KW environmental pollutant; cellular response; gene expression profile;
 KW toxic response; liver necrosis; fatty liver disease;
 KW protein adduct formation; hepatitis; dog; ds.
 XX Canis familiaris.
 XX WO2004063324-A2.
 XX 29-JUL-2004.
 XX 05-MAY-2003; 2003WO-US013853.
 XX 03-MAY-2002; 2002US-0377240P.
 XX (GENE-) GENE LOGIC INC.
 XX (PFIZ) PFIZER PROD INC.
 XX Diggins JC, Porter M, Wei T;
 XX WPI; 2004-561890/54.
 XX New isolated nucleic acid molecule, useful for drug screening and
 PT toxicity assays or for assessing the impact, including toxicity, of a
 PT compound, pharmaceutical agent or environmental pollutant on a cell or
 PT living organism.
 XX Claim 1; SEQ ID NO 8126; 41pp; English.
 XX This invention is related to a novel isolated canine nucleic acid
 CC sequences and the construction of canine microarrays containing a
 CC significant portion of the canine genome. The isolated canine nucleic
 CC acid sequences of the invention may be useful for drug screening and
 CC toxicity assays. The invention is therefore useful for assessing the
 CC impact, including toxicity, of a compound, pharmaceutical agent or
 CC environmental pollutant on a cell or living organism. The methods are
 CC useful for detecting genes that are up- or down-regulated in canines in a
 CC disease state. The sequences are useful as diagnostic agents or markers
 CC to detect a cellular response in a sample individually or as part of a
 CC gene expression profile. It is also useful as a target for agents that
 CC modulate gene expression or activity. The database is useful for the cell
 CC producing electronic Northernblots that allow the user to determine the cell
 CC type or tissue in which a given gene is expressed and to allow
 CC determination of the abundance or expression level of a given gene in a

CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues, to identify their cognate ligands
CC or binding partners, and in chemotaxis, and can be used as a food
CC additive or preservative to modify storage properties. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above, and in diagnostic
CC immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay
CC (ELISA). The present sequence represents a human secreted protein-
CC encoding cDNA of the invention
XX
XX Sequence 915 BP; 209 A; 253 C; 247 G; 202 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 5.23e-76 Length: 915
Score: 809.00 Matches: 155
Percent Similarity: 96.88% Conservative: 0
Best Local Similarity: 96.88% Mismatches: 5
Query Match: 40.88% Indels: 0
DB: Gaps: 0

US-09-774-954-2 (1-365) x AAD07629 (1-915)

QY 206 HisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrpLysAsnAla 225
DB 16 CACTTTGCCCGAACCTACCGCGGCAATTCATCTGGCATTTGGCTCTGACTGGAAAGACGCC 75
QY 226 CysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerProGlnCysVal 245
DB 76 TGTGCCATGCTGAAGGACGGGAGCTGCAGGCTCGCCTTCATGGCCTCTCCGCGAGTGTGTG 135
QY 246 GlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuProAspLeuLys 265
DB 136 GGTCTACAGCCGACGACAGCGGCGCCCTCACGATGACTATGTGCTGCTGACCTGAAG 195
QY 266 GluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAlaGlnSerValTyrVal 285
DB 196 GAGATCCAGAGGCGTGTGAAGCTCTGGGTGAGTGTGGATGCCAGTCCGCTCAGCTT 255
QY 286 AlaThrAspSerGluSerTyrValProGluLeuGlnLeuPheLysGlyLysValLys 305
DB 256 GCTACTGATTCGAGAGTATGTGCTGTGAGCTCCACAGCTCTTCAAGGGAAGGTGAAG 315
QY 306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
DB 316 GTGGTGAGCTGAAGCTGAGGTGGCCAGTCTGACCTGTACATCTCGGCCAAGCGGAC 375
QY 326 HisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeu 345
DB 376 CACTTTATGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 435
QY 346 GlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArgAspGluPhe 365
DB 436 CAGGGGAGSCCGTCT 495

RESULT 11

ABL13946/c
ID ABL13946 standard; cDNA; 3264 BP.

XX
AC ABL13946;

XX
DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;

XX
KW pharmaceutical; gene; ss.

XX
OS Drosophila melanogaster.

XX
FN WO200171042-A2.

XX
PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB69843.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.16e-74 Length: 3264
Score: 796.00 Matches: 167
Percent Similarity: 58.27% Conservative: 62
Best Local Similarity: 42.49% Mismatches: 119
Query Match: 40.22% Indels: 45
DB: Gaps: 10

US-09-774-954-2 (1-365) x ABL13946 (1-3264)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
DB 2183 GATCCCAATGGCTACCTCCTACTCTCGTGTATGGTAAAGTTTCATCTGGAGCAACC 2124
QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
DB 2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCGCCGCCACTT 2064
QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProTrpI1 48
DB 2063 CTTGGGATCATTTGGCTTCCCAAGGCGCTTATCGACCTGATCTGCGCGGTGGT 2004
QY 48 eGluTyrGlnHisHisLysProPheThrAsnLeuHisValSerTyrGlnLysTyrPh 68
DB 2003 GGAGTATCGT-----AGGGGTGAACCTGCATCCGTCAGTACCGTTCAACACATATT 1950
QY 68 eLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMetGluLy 88
DB 1949 TGAAGTGGAGCCCTTGAAGGAATACCATCCGTCATCACCATTGGCAGATTTCATGTGCA 1890
QY 88 sLeuAlaProThrHisTrpProGluLysArgValAlaTyrCysPheGlu----- 105
DB 1889 CTTGGCGCAGCATTTTGGCCAGAAATCGGAGCGAGTGTCTATTTGTCTACAAGGAACGATA 1830
QY 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123
DB 1829 TAGCCTTCAGCAGGAGAAGAACGATCCAGACAAAGCCCAATTGGCCACCGCAGGATGCA 1770
QY 123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
DB 1769 TCCTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTTGTGCGGTGAGAGTTCTATGC 1710

Db 141 TGCAGATCATATCCCGTACTCGGTTTCCCGAGCTCCGGCTAGTTTCTGTTTCAGCT 82
QY 180 uGluHisArgProLeuGlnLysTyrMetValTrrSer-aspGluMetValLysThrGlyG 200
Db 81 AGAGAACTGCAAGCTGCGGCTACTTGCAGTGGAGTCAACGGT----- 38
QY 200 luAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlyS 220
Db 37 -----ATAGGGAAGCATCTAAGGATTTCAT 13
QY 220 erAsp 221
Db 12 CCGAG 8
RESULT 13
AAK79915
ID AAK79915 standard; DNA; 10331 BP.
XX
AC AAK79915;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239315P.
PR 13-OCT-2000; 2000US-0239377P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.

PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	03-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-02311242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-02311243P.	PR	17-NOV-2000;	2000US-0249246P.
PR	08-SEP-2000;	2000US-02311244P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-02311413P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-02311414P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250160P.
PR	12-SEP-2000;	2000US-0232080P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232399P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234223P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234997P.	XX		
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	XX		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	XX		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	XX		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	XX		
PR	13-OCT-2000;	2000US-0239935P.	XX		
PR	13-OCT-2000;	2000US-0239937P.	XX		
PR	20-OCT-2000;	2000US-0240960P.	XX		
PR	20-OCT-2000;	2000US-0241221P.	XX		
PR	20-OCT-2000;	2000US-0241785P.	XX		
PR	20-OCT-2000;	2000US-0241786P.	XX		
PR	20-OCT-2000;	2000US-0241787P.	XX		
PR	20-OCT-2000;	2000US-0241808P.	XX		
PR	20-OCT-2000;	2000US-0241809P.	XX		
PR	20-OCT-2000;	2000US-0241826P.	XX		
PR	01-NOV-2000;	2000US-0244617P.	XX		
PR	08-NOV-2000;	2000US-0246474P.	XX		
PR	08-NOV-2000;	2000US-0246475P.	XX		
PR	08-NOV-2000;	2000US-0246476P.	XX		
PR	08-NOV-2000;	2000US-0246477P.	XX		
PR	08-NOV-2000;	2000US-0246478P.	XX		
PR	08-NOV-2000;	2000US-0246523P.	XX		
PR	08-NOV-2000;	2000US-0246524P.	XX		
PR	08-NOV-2000;	2000US-0246525P.	XX		
PR	08-NOV-2000;	2000US-0246526P.	XX		
PR	08-NOV-2000;	2000US-0246527P.	XX		
PR	08-NOV-2000;	2000US-0246528P.	XX		
PR	08-NOV-2000;	2000US-0246532P.	XX		
PR	08-NOV-2000;	2000US-0246609P.	XX		
PR	08-NOV-2000;	2000US-0246610P.	XX		
PR	08-NOV-2000;	2000US-0246611P.	XX		
PR	08-NOV-2000;	2000US-0246613P.	XX		
PR	17-NOV-2000;	2000US-0249207P.	XX		
PR	17-NOV-2000;	2000US-0249208P.	XX		
PR	17-NOV-2000;	2000US-0249209P.	XX		
PR	17-NOV-2000;	2000US-0249210P.	XX		
PR	17-NOV-2000;	2000US-0249211P.	XX		
PR	17-NOV-2000;	2000US-0249212P.	XX		
PR	17-NOV-2000;	2000US-0249213P.	XX		

US-09-774-954-2 (1-365) x AAL03344 (1-10331)

Qy	223	LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro	242
Db	2491	ARGAACCCCTGTGCCATGCTGAGAGCGGACTGCAGGCTCGCACTTCATGCGCTCTCCG	2550
Qy	243	GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro	262
Db	2551	CAGTGTGTGGGTACAGCGCAGCACAGCGGCCCCCTCACGATGACTATGTGCTGCCT	2610
Qy	263	AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer	282
Db	2611	GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCCAGTCG	2670
Qy	283	ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly	302
Db	2671	GTCCTACGTTGCTACTGATTCCGAGATTATGTGCTGAGCTCCACAGCTCTTCAAGGG	2730
Qy	303	LysValLysVal	306

Alignment Scores:

Pred. No.:	1.76e-33	Length:	10331
Score:	424.00	Matches:	83
Percent Similarity:	98.81%	Conservative:	0
Best Local Similarity:	98.81%	Mismatches:	1
Query Match:	21.42%	Indels:	0
DB:	4	Gaps:	0

Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Thu Oct 27 11:17:25 2005

Db 2731 AAGGTATGTGTG 2742
RESULT 15
ACH32225
ID ACH32225 standard; cDNA; 479 BP.
XX
AC ACH32225;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human endothelial cell cDNA #358.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 19437; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX for chromosome and gene mapping, in the recombinant production of
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;

Search completed: October 26, 2005, 09:51:26
Job time : 1056.3 secs

Alignment Scores:
Pred. No.: 3 9e-32 Length: 479
Score: 393.00 Matches: 76
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.86% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-2 (1-365) x ACH32225 (1-479)
QY 290 GluSerTyrValProGluLeuGlnLeuPheLysGlyLysValValSerLeu 309

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 ; Search time 1559.68 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLCPMGR.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9772377 seqs, 4126317084 residues
Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat_25102005_105434_6433/app_query.fasta_1.917
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPC1=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09774954 @Cgn 1 1 684 @runat_25102005_105434_6433
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
25: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
27: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	1100	11	US-09-774-954-16
2	1979	100.0	1300	11	US-09-774-954-1
3	1979	100.0	1514	11	US-09-774-954-1
4	1979	100.0	5218	16	US-10-301-822-168
5	1979	100.0	5218	22	US-10-956-157-4752
6	1979	100.0	11284	11	US-09-774-954-5
7	1844	93.2	5009	11	US-09-774-954-7
8	889	44.9	1400	22	US-10-956-157-9986
9	889	44.9	1506	22	US-10-956-157-4751
10	831	42.0	1209	26	US-11-097-143-18161
11	796	40.2	3264	26	US-11-097-143-18160
12	506.5	25.6	3793	26	US-11-097-143-18082
13	424	21.4	10331	10	US-09-764-891-6032
14	393	19.9	479	10	US-09-918-995-19437
15	332	19.8	3567	26	US-11-097-143-18169
16	353	17.8	477	10	US-09-918-995-22744
17	264	13.3	402	9	US-09-783-590-11501
18	164.5	8.3	1967	20	US-10-471-450-52
19	164	8.3	1926	26	US-11-097-143-36191
20	164	8.3	3926	26	US-11-097-143-36190
21	142	7.2	2434	20	US-10-471-450-34
22	120	6.1	521	13	US-09-925-065A-602444
23	120	6.1	760	24	US-10-820-474A-236
24	117	5.9	2238	19	US-10-424-599-2236
25	115.5	5.8	4104	26	US-11-097-143-36187
26	110	5.6	1574	22	US-10-764-420-2104
27	109	5.5	1551	18	US-10-062-674-1876
28	108	5.5	3052	22	US-10-956-157-41465
29	106.5	5.4	2166	20	US-10-437-963-59638
30	105.5	5.3	1512	22	US-10-956-157-1986
31	103.5	5.2	960	18	US-10-309-290-119
32	103.5	5.2	960	18	US-10-425-114-31610
33	103	5.2	2122	21	US-10-425-115-143813
34	103	5.2	2122	21	US-10-128-714-2312
35	102.5	5.2	783	16	US-10-128-714-2312
36	102.5	5.2	907	16	US-10-128-714-312
37	102.5	5.2	2907	16	US-10-128-714-312
38	102	5.2	2790	20	US-10-437-963-71433
39	102	5.2	29069	24	US-10-915-740A-44
40	102	5.2	2242716	24	US-10-915-740A-1068
41	101	5.1	1617	20	US-10-437-963-11436
42	101	5.1	2600	24	US-10-450-763-10656
43	100.5	5.1	24678	22	US-10-741-600-17702
44	99	5.0	14399	19	US-10-287-226-321
45	99	5.0	140152	21	US-10-684-422-66

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-NOV-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1100 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16
Alignment Scores:
Pred. No.: 1,276-238 Length: 1100
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0
US-09-774-954-2 (1-365) x US-09-774-954-16 (1-1100)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCCCGCGGGCTCTGGGACCCCGCGGTACCTGCTCTACTGCTCCCTGTCATGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArg 40
Db 61 TTGGGAACACGCGCATCACTTCTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db 121 ACCITGGCTGTCCCTCTCTGGATTGAGTACCAGCATCACAAAGCTCTTTCACCAACCTC 180
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValile 80
Db 181 CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
Db 241 AGCTTGGAGGATTTCATGGAGAGCTGGACCCACCCACCTGGCCCTTGAGNAGCGGGTG 300
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysThrCysProMetLys 120
Db 301 GCATACTGCTTTGAGTGGCAGCCAGCCAGCCAGCATAGAGAGCGTGGCCCATGAAG 360
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 361 GAAGGAAACCCCTTTGGCCCATCTTGGGATCATGTTTCTGAGTTTCAACAAGTCGGAG 420
QY 141 LeuPheThrGlyLysLeuSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 421 CTTTTCACAGCAATTCCTTTCAGTGCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 480

QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db 481 CCAAGAGACATCCGGTGTCTGCTCCAGGAGGCCAGCCAGTTCCTCCCGTCTAGAA 540
QY 181 GluHisArgProLeuGlnLysTyrMetValTyrSerAspGluMetValLysThrGlyGlu 200
Db 541 GAACACAGGCCACTACAGAGTACATGGTATGGTACAGGAAATGGTGAAGACGGGAG 600
QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 601 GCCCAGATTATGCCCCCTTGTCCGGCCCTATGTGGGCATTCATCTGGCATTTGGCTCT 660
QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 661 GACTGGAAGAACGCTGTGCCATCTGAGAGACGGGACTGCGAGGCTGCGACTTCATGGCC 720
QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 721 TCTCCGAGTGTGGGCTACAGCCGCGAGCACAGCGGCCCTCCAGATGACTATGTGC 780
QY 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAla 280
Db 781 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGCTCGCTGGATGCC 840
QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 841 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACACAGCTCTTC 900
QY 301 LysGlyLysValLysValLysSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 901 AAGGGGAAGGTGAAGGTGGTGAAGCTTGAAGCTTGAAGTGGCCAGGTCGACCTGTACATC 960
QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 961 CTCGCCCAAGCCAGCACCTTATTTGGCAACTGTGTCTCTCTCTCCTGCTGCTTGTGAAG 1020
QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
Db 1021 CGGAGCGGACCTCCAGGGAGGCGCTCTTCTTCTTGGCATGACAGGCGCCCTTAAG 1080

QY 361 LeuArgAspGluPhe 365
Db 1081 CTGCGGAGCAGGTTTC 1095

RESULT 2
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
;

ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.
 REGISTRATION NUMBER: 39,044
 REFERENCE/DOCKET NUMBER: P1041P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-1489
 TELEFAX: 650/952-9881
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1300 base pairs
 TYPE: Nucleic Acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Alignment Scores:
 Pred. No.: 1-63e-238 Length: 1300
 Score: 1979.00 Matches: 365
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 11 Gaps: 0

US-09-774-954-2 (1-365) x US-09-774-954-4 (1-1300)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 DB 136 ATCCCGCGGGCTCTGGGACCGCGGGTACCTCTCTACTGCGCCCTGCATGGGGCGC 195
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeuLeuLeuLeu 40
 DB 196 TTTGGGAACAGCGCCGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 255
 QY 41 ThrLeuAlaValProProTrpPilleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
 DB 256 ACCTTGGCTCTCTCTGGATTGAGTACCAAGCTCCTCTTCAACCAACCTC 315
 QY 61 HisValSerTyrGlnLysTyrPheLeuLeuGluProLeuGlnAlaTyrHisArgValile 80
 DB 316 CATGTGCTCTTACCAAGATGATTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 375
 QY 81 SerLeuGluAspPheMetGluLeuLeuAlaProThrHisTrpProGluLysArgVal 100
 DB 376 AGCTTGGAGATTTCATGGAGAAGCTGGACCCACCCACTGGCCCTGAGAGCGGGTG 435
 QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 DB 436 GCATCTGCTTTGAGGTGGCAGCCAGCCAGCAAGCCAGATGAAGAGCTGCCCATGAAG 495
 QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
 DB 496 GAAGGAACCCCTTTGGCCCATCTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555
 QY 141 LeuPheThrGlyLeuSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
 DB 556 CTTTATACAGCATTTCTCTTCAAGTCTTCTTACAGAAACATGGAGCCAGAGATTTCCT 615
 QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
 DB 616 CCAAGGAACATCCGGTGTCTTGGCTGCCAGGAGCCCGCCAGTTCCCGCTCCTAGAG 675
 QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 DB 676 GAACACAGGCCACTACAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 735
 QY 201 AlaGlnLeuHisAlaHisLeuValArgProTyrValGlyLeuHisLeuArgLleGlySer 220
 DB 736 GCCCAGATTTCATGCCCATCTTGTCCGCCCTTATGTGGGCTATCATCTGCGGATTTGGCTCT 795
 QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 DB 796 GACTGGAGAGACGCTGTGGCATCTGTGAGGACGGGACTGCAGGCTCGCATTCTCATGGCC 855

QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
 DB 856 TCTCCGAGTGTGTGGGTACAGCGCAGCACAGCGGCCCTCAGATGACTATGTGC 915
 QY 261 LeuProAspLeuLysGluLeuGlnArgAlaValLysLeuTyrValArgSerLeuAspAla 280
 DB 916 CTGCCTGACCTGAAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGGCTGGATGCC 975
 QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
 DB 976 CAGTGGTCTACGTGCTACTGATTCGAGAGTTATGCTGCTGAGCTCCAGAGCTCTTC 1035
 QY 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
 DB 1036 AAAGGAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCAGCTCGACCTGTATCATC 1095
 QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
 DB 1096 CTCGCGCAAGCGACCACTTTATTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCT 1155
 QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
 DB 1156 CGGGAGCGGAGCTCCAGGGGAGCGGCTCTTTCTTCTGCGCATGGACAGGCCCTTAA 1215
 QY 361 LeuArgAspGluPhe 365
 DB 1216 CTGCGGAGCAGTTC 1230

RESULT 3

US-09-774-954-1
 ; Sequence 1, Application US/09774954
 ; Publication No. US20040241645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Winpatin (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/774,954
 ; FILING DATE: 30-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,741
 ; FILING DATE: 26-Nov-1997
 ; APPLICATION NUMBER: 08/792,498
 ; FILING DATE: 31-Jan-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1041P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1514 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 ; US-09-774-954-1

Alignment Scores:		2,05e-238	Length: 1514
Pred. No.:	Score:	1979.00	Matches: 365
Percent Similarity:		100.00%	Conservative: 0
Best Local Similarity:		100.00%	Mismatches: 0
Query Match:		100.00%	Indels: 0
DB:		11	Gaps: 0
US-09-774-954-2 (1-365) x US-09-774-954-1 (1-1514)			
QY	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	1	ATGCCCGCGGGCTCTCGGAGCCCGCGGTACTCTACTGCGCCCTGCATGGGGCGC	60
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	61	TTTGGGAACCGCCGATCACTCTGGGCTCTGTGGCAATTGGCAAGCTGTAAACCGT	120
QY	41	ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu	60
DB	121	ACCTTGGCTGCTCCCTCTGTGATTGAGTACAGCATCAAGCCCTCTTTTCAACAACCTC	180
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle	80
DB	181	CATGTGCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC	240
QY	81	SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal	100
DB	241	AGCTTGGAGGATTTTCATGGAGAGCTGGCACCACCTGCGCCCTTGAGAGCGGGTG	300
QY	101	AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys	120
DB	301	GCATACTGCTTTGAGTGGGAGCCAGCCAGGAGCCAGATTAAGAGACGTGCCCATGAG	360
QY	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140
DB	361	GAAGGAACCCCTTTGGGCCATTTCTGGATCATGTTTCAATGAGTTTCAACAAGTCGAG	420
QY	141	LeuPheThrGlyLysPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer	160
DB	421	CTTTTACAGGCAATTCCTTCAGTGTCTTCTACAGAGACACATGGAGCCAGAGATTTC	480
QY	161	ProLysGluHisProValLeuAlaProGlyAlaProAlaGlnPheProValLeuGlu	180
DB	481	CCAAAGGAACATCCGGTGTGGCCCTGCCAGGAGCCCGAGCCAGTTCCTCTTGAA	540
QY	181	GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu	200
DB	541	GAACACAGGCCACTACAGAGTACATGGTATGCTCAGACGAAATGGTGAAGCGGAGAG	600
QY	201	AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer	220
DB	601	GCCAGATTTCATGCCCCACCTTGTCCGGCCCTATGTGGCAATTCATCTGCGCATTTGG	660
QY	221	AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
DB	661	GACTGGAGAACCCCTGTGCTCATCTGAGAGCGGACTGCGAGGCTCCCATTTGATGGCC	720
QY	241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
DB	721	TCTCCGAGTGTGTGGGCTACAGCGGAGCAGCAGCGCCCTCCATGACATATGTGC	780
QY	261	LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla	280
DB	781	CTGCCCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC	840
QY	281	GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe	300
DB	841	CAGTGGCTTACGTTGCTACTGATTCGAGAGTATGTGCTGAGCTCCACACAGCTCTTC	900
QY	301	LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle	320
DB	901	AAAGGAAGGTGAAGGTGCTGAGCCTGAAGCCTGAGTGGCCAGGTGACCTGTATCATC	960
QY	321	LeuGlyClnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys	340
DB	961	CTGCGCAAGCCGACCACTTTATGGCAACTGTGTCTCTCTTCCTCCTGCTTGTGAAG	1020
QY	341	ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys	360
DB	1021	CGGAGCGGAGACCTCCAGGGGAGCCGCTTCTTCTTCTGCGCATGGACAGCCCTTAAG	1080
QY	361	LeuArgAspGluPhe	365
DB	1081	CTGCGGAGCAGGTTC	1095
RESULT 4			
US-10-301-822-168			
; Sequence 168, Application US/10301822			
; Publication No. US20030148410A1			
; GENERAL INFORMATION:			
; APPLICANT: Millennium Pharmaceuticals, Inc.			
; APPLICANT: Berger, Allison			
; APPLICANT: Guillemette, Tracy L.			
; APPLICANT: Kamatkar, Shubhangi			
; APPLICANT: Schlegel, Robert			
; APPLICANT: Monahan, John E.			
; APPLICANT: Thibodeau, Stephen N.			
; APPLICANT: Burgart, Lawrence J.			
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND			
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND			
; TITLE OF INVENTION: THERAPY OF COLON CANCER			
; FILE REFERENCE: MEM01-029P2RNM			
; CURRENT APPLICATION NUMBER: US/10/301,822			
; CURRENT FILING DATE: 2002-11-21			
; PRIOR APPLICATION NUMBER: US 60/339,971			
; PRIOR FILING DATE: 2001-12-10			
; PRIOR APPLICATION NUMBER: US 60/361,978			
; PRIOR FILING DATE: 2002-03-05			
; PRIOR APPLICATION NUMBER: US 60/381,988			
; PRIOR FILING DATE: 2002-05-20			
; NUMBER OF SEQ ID NOS: 228			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 168			
; LENGTH: 5218			
; TYPE: DNA			
; ORGANISM: Homo Sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (50)...(1216)			
US-10-301-822-168			
Alignment Scores:			
Pred. No.:	1,31e-237	Length: 5218	
Score:	1979.00	Matches: 365	
Percent Similarity:	100.00%	Conservative: 0	
Best Local Similarity:	100.00%	Mismatches: 0	
Query Match:	100.00%	Indels: 0	
DB:	16	Gaps: 0	
US-09-774-954-2 (1-365) x US-10-301-822-168 (1-5218)			
QY	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	119	ATGCTCGGGGCTCTGGGAGCCCGCGGTACTCTTACTGCCCTGCATGGGGGC	178
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	179	TTTGGGAACCAAGCCGATCACTCTTGGGCTCTTGGCAATTGCAAGCTGTAAACCGT	238
QY	41	ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu	60
DB	239	ACCTTGGCTGTCTCTCTGGATTGATTGATACCAAGCATCACAAGCTCTCTTTCACCAACCTC	298
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle	80

```

Db 299 CATGTGCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 358
Qy 81 SerLeuGluaspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
Db 359 AGCTTGAGGATTTTCATGTGAGAAAGCTGGACCCACCACTGGCCCTGAGAGCGGGTG 418
Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 419 GCATACTGCTTTGAGGTGGAGCCGAGGAGCCAGATAGAGAGCGTGCCCATGAAG 478
Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 479 GAAGGAAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGATGTTCAACAAGTCGGAG 538
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 539 CTTTTCACAGGCATTTCTCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 598
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyValAlaProAlaGlnPheProValLeuGlu 180
Db 599 CCAAGGAAACATCCGGTGCTTGGCCCTGCGAGGAGCCCGCCAGCTTCCCGCTCTAGAG 658
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 659 GAACACAGGCCACTTACAGAGATCATGTGATGTGTCAGACGGAATGGTGAAGCGGGAGAG 718
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 719 GCCCAGATTTCATGCCACCTTTGTCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 778
Qy 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 779 GACTTGGAGAACCGCTGTGCTGTCATGCTGAAGAGCGGACTCGAGCTCGCACTTCATGGCC 838
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 839 TCTCCGAGTGTGTGGCTACAGCGGACGACAGCGGCCCGCCCTCAGATGACTATGTGC 898
Qy 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db 899 CTGCCTCACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGCC 958
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 959 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTTC 1018
Qy 301 LysGlyLysValLysValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 1019 AAAGGGAAGGTGAAGTGTGAGCTGAGCCTGAGCCTGAGGTGGCCAGGTTCGACCTGTACATC 1078
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 1079 CTCGGCCAAAGCCGACCACTTATTGGCAACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1138
Qy 341 ArgGluArgAspLeuGlnArgProSerSerPheGlyMetAspArgProProLys 360
Db 1139 CGGAGCGGAGCACTTCCAGGAGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1198
Qy 361 LeuArgAspGluPhe 365
Db 1199 CTGCGGACGAGTTTC 1213

RESULT 5
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
```

```

; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
; LENGTH: 5218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores:
Pred. No.: 1,31e-237 Length: 5218
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-2 (1-365) x US-10-956-157-4752 (1-5218)

Qy 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 119 ATGCCTCGGGCTCTCTGGACCGCGCGGTACTCTCTACTGCCCCCTGCATGGGCGC 178
Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 179 TTTGGGAACACAGGCCGATCATCTTCTTGGGCTCTCTGGCATTTTCAAAGCTGTCTAAACCGT 238
Qy 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProPheThrAsnLeu 60
Db 239 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCACAAAGCTCTCTTTCACCAACCTC 298
Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db 299 CATGTGCTTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 358
Qy 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
Db 359 AGCTTGAGGATTTTCATGGAGAGCTGGCACCCACCACTGGCCCCCTGAGAAGCGGGTG 418
Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 419 GCATACTGCTTTGAGGTGGAGCCGAGGAGCCAGATAGAGAGCGTGCCTCATGAAG 478
Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 479 GAAGGAAACCCCTTTGGCCCATTTCTGGATCAGTTTCATGTGATGTTCAACAAGTCGGAG 538
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 539 CTTTTCACAGGCATTTCTCTTCAGTGTCTTCTACAGAGAAACAATGGAGCCAGAGATTTTCT 598
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyValAlaProAlaGlnPheProValLeuGlu 180
Db 599 CCAAGGAAACATCCGGTGCTTGGCCCTGCGAGGAGCCCGCCAGCTTCCCGCTCTAGAG 658
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 659 GAACACAGGCCACTTACAGAGATCATGTGATGTGTCAGACGGAATGGTGAAGCGGGAGAG 718
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 719 GCCCAGATTTCATGCCACCTTTGTCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 778
Qy 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 779 GACTTGGAGAACCGCTGTGCTGTCATGCTGAAGAGCGGACTCGAGCTCGCACTTCATGGCC 838
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 839 TCTCCGAGTGTGTGGCTACAGCGGACGACAGCGGCCCGCCCTCAGATGACTATGTGC 898
Qy 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
```

```

Db      899  CTGCTGACCTGAGGAGATCCAGGGCTGTGAAGCTCTGGGTGAGTGGTGGCC 958
Qy      281  GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db      959  CAGTGGTCTAGTGTCTACTGATTCGAGAGTATGTGCTGAGCTCCAAAGCTCTTC 1018
Qy      301  LysGlyLysValValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db      1019  AAAGGGAAGGTGAAGGTGTGAGCTGAAGCCTGAGGTGGCCAGGTCGACCTGTACATC 1078
Qy      321  LeuGlyGlnAlaAspPheLeGlyAsnCysValSerSerPheThrAlaPheValIys 340
Db      1079  CTGGGCCAAGCGACCACTTATTGGCACTGTGTCTCTCTCACTGGCTTTGTGAAG 1138
Qy      341  ArgGluArgAspLeuGlnGlyValGProSerPhePheGlyMetAspArgProProlys 360
Db      1139  CGGAGCGGGACCTCCAGGGAGGCGCTCTCTCTTCCTGGCATGGACAGGCCCTTAG 1198
Qy      361  LeuArgAspGluPhe 365
Db      1199  CTGGGGACGAGTTC 1213

```

RESULT 6

```

US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; LENGTH: 11284 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

Alignment Scores:
Pred. No.: 4,19e-237 Length: 11284
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

```

Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0
US-09-774-954-2 (1-365) x US-09-774-954-5 (1-11284)
Qy      1  MetProAlaGlySerTrpAspProAlaGlyTyrLeuTyrCysProCysMetGlyArg 20
Db      4236  ATCCCGCGGGCTCTCTGGAGCCCGGCGGGTACCTGCTACTGCCCCCTGCATGGGGCGC 4295
Qy      21  PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      4296  TTTGGGAACACGCGCGATCACTCTCTGGGCTCTCTGGCATTTGCNAGCTGCTAACCCT 4355
Qy      41  ThrLeuAlaValProProTrpIleGluTyrGlnHisLysLysProProPheThrAsnLeu 60
Db      4356  ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCTACCAAGCTCTCTTCCACCACCTC 4415
Qy      61  HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db      4416  CATGTGTCTCTACAGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 4475
Qy      81  SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
Db      4476  AGCTTGGAGGATTTTCATGGAGAAGCTGGCACCCACCACCTGGCCCTCTGAGAGCGGGTG 4535
Qy      101  AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysLysThrCysProMetLys 120
Db      4536  GCATAGTCTTTGAGGTGGCAGCCCGGAGCCAGATAAGAGACGTCGCCCATGAAG 4595
Qy      121  GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db      4596  GAAGGAACCCCTTTGGCCCACTTCTGGGATCACTTCACTGAGTTTCAACAAGTCGGAG 4655
Qy      141  LeuPheThrGlyLysPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db      4656  CTTTTTACAGGCAATTTCTTTCAGTGTCTTCTCAGAGAACAATGGAGCCAGAGATTTCT 4715
Qy      161  ProlysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db      4716  CCAAAGCAACATCCGCTGCTTGGCCCTCTCCAGGAGCCCGCCAGCTCCCGCTCTAGAG 4775
Qy      181  GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db      4776  GNACACAGGCCACTACAGAAAGTACATGTATGTGTGAGCAAAATGCTGAAGACGGGAGAG 4835
Qy      201  AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db      4836  GCCCAGATTATGCCCCACCTTGTCCGGCCCTATGTGGGCATTTCACTGGCATTTGGCTCT 4895
Qy      221  AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db      4896  GACTGGGAAGAACCCCTGTGCCATGCTGNAAGACGGGACTGCAGGCTCGCACTTCATGGCC 4955
Qy      241  SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db      4956  TCTCCGAGTGTGTGGGCTACAGCCGACACAGCGGCCCTCTACGATGACTATGTGC 5015
Qy      261  LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAla 280
Db      5016  CTGCTGACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCCGTGATGCC 5075
Qy      281  GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db      5076  CAGTCCGCTCTACGTTCTGACTGATTCGAGAGTTATGTGCTGAGCTCCAAAGCTCTTC 5135
Qy      301  LysGlyLysValLysValValSerLeuLysProLysValAlaGlnValAspLeuTyrIle 320
Db      5136  AAAGGGAAGGTGAAGGTGTGAGCTTGAAGCCTGAAGCCTGAGGTGGCCAGGCTCGACCTGTACATC 5195
Qy      321  LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db      5196  CTCGGCCCAAGCCGACCACTTATTATGGCAACTGTGTCTCTCTCTTCACTGCTTTGTGAAG 5255

```

QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
Db 5256 CCGGAGCGGACCTCCAGGGAGGCGCTCTCTTTCTTGGCATGACACAGGCCCTTAAG 5315
QY 361 LeuArgAspGluPhe 365
Db 5316 CTGCGGACGAGTTC 5330

RESULT 7

US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; TYPE: Nucleic Acid
; LENGTH: 5009 base pairs
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

Alignment Scores:
Pred. No.: 1.32e-220 Length: 5009
Score: 1844.00 Matches: 343
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 93.18% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-2 (1-365) x US-09-774-954-7 (1-5009)

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
Db 2 AACGAGCGCGATCACTCTTGGGCTCTGSCATTTGCAAGCTGTAAACCGTACCTTG 61
QY 43 AlaValProProThrIleGluTyrGlnHisLysProPheThrAsnLeuHisVal 62
Db 62 GCTGTCCCTCTTGGATTGAGTACAGCATCACAGGCTCTTTTCCACCAACCTCATGTG 121
QY 63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82

Db 122 TCCTACCAGAGTACTTCAAGCTGCAGCCCTCCAGGCTTACCATCGGTGCATCACTGTTG 181
QY 83 GluAspPheMetGluLysLeuAlaProThrHisTyrProProGluLysArgValAlaTyr 102
Db 182 GAGGATTTTCATGGAGAGCTGGACCCCACTGGCCCCCTGAGAGGGGGTGGCATAC 241
QY 103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
Db 242 TGTCTTGGTGGCAGCCAGCCAGCCAGATAGAGACGCTGCCCATGAAGGAAGGA 301
QY 123 AsnProPheGlyProPheThrAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
Db 302 AACCCCTTTGGCCCATTCCTGGGATTCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTT 361
QY 143 ThrGlyLysSerPheSerAlaSerTyrArgGluGlnTyrSerGlnArgPheSerProLys 162
Db 362 ACAGGCATTTCTTTCAGTCTTCTTACAGAGAACATGAGCCAGAGATTTTCTCCAAG 421
QY 163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
Db 422 GAACATCCGCTGCTTGCCTGCCAGGAGCCCGCCAGTTCCTCCCTAGAGGAACAC 481
QY 183 ArgProLeuGlnLysTyrMetValTyrSerAspGluMetValLysThrGlyGluAlaGln 202
Db 482 AGGCCACTACAGAAGTACATGTTGTTGTCAGAGAAATGTTGAAGACGGGAGGCCCCAG 541
QY 203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTyr 222
Db 542 ATTCTATGCCACCTTGTCCGCCCTATGTGGGATTTCACTGGGCATTTGGCTGACTGG 601
QY 223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
Db 602 AAGAAGCCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACCTCATGGCTCTCCG 661
QY 243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
Db 662 CAGTGTGGGCTACAGCGCGACACAGCGGCCCTCCAGATGACTATGTGCTGCTGCCT 721
QY 263 AspLeuLysGluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAlaGlnSer 282
Db 722 GACCTGAAGGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCAGTCG 781
QY 283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
Db 782 GTCTAGTGTCTACTGATTCGAGAGTTAIGTSCCTGAGCTCCACACAGCTCTTCAAAGG 841
QY 303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
Db 842 AAGGTGAAGGTGCTGAGCCCTGAAGCCCTGAGGTGGCCCGGCTGATCATCTCTCGC 901
QY 323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
Db 902 CAAGCCGACCATTTATTTGGCAACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 961
QY 343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
Db 962 CCGGACCTCCAGGGAGGCGCTCTTCTTCTTGGCATGAGACGGCCCCCTTAAGTCGG 1021
QY 363 AspGluPhe 365
Db 1022 GACGAGTTC 1030

RESULT 8

US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157

; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-9986

Alignment Scores:
Pred. No.: 5,67e-101 Length: 1400
Score: 889.00 Matches: 158
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 44.92% Gaps: 0

US-09-774-954-2 (1-365) x US-10-956-157-9986 (1-1400)

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 13 ATGCTGCGGGCTCTGGGACCCGCGGTACCTGCTACTGCTCCCTCATGGGGCGC 72
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 73 TTGGGAAACCGCCGATCACTTCTGGGCTCTCGGCATTTCGAAAGCTGCTAAACCGT 132
QY 41 ThrLeuAlaValProProTTPileGluTyrGlnHisHisLysProProPheThrAsnLeu 60
DB 133 ACCTTGCTGCTCCCTTCCTTGGATTGAGTACCAAGCTCAGCCCTCCAGGCTTACCATCGGGTCATC 192
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
DB 193 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 252
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTTPProGluLysArgVal 100
DB 253 AGCTTGGAGGATTCATGGAGAAGCTGGCACCCACCACTGGGCCCTCGAGAAGCGGGTG 312
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
DB 313 GCATCTGCTTGGAGTGGAGCCAGCCAGCCAGATAGAGAAGCTGCCCATGAG 372
QY 121 GluGlyAsnProPheGlyProPheTTPAspGlnPheHisValSerPheAsnLysSerGlu 140
DB 373 GAAGGAAACCCCTTTGGGCCATTCCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 432
QY 141 LeuPheThrGlyLeSerPheSerAlaSerTyrArgGluGlnTTPSerGlnArg 158
DB 433 CTTTTTACAGGCAATTCCTTCAGTGTCTCTACAGAGAACAATGGAGCCAGAG 486

RESULT 9

US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:
Pred. No.: 6.33e-101 Length: 1506

Score: 889.00 Matches: 158
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 44.92% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-2 (1-365) x US-10-956-157-4751 (1-1506)

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 119 ATGCTGCGGGCTCTGGGACCCGCGGTACCTGCTACTGCTCCCTCATGGGGCGC 178
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 179 TTGGGAAACCGCCGATCACTTCTGGGCTCTCGGCATTTCGAAAGCTGCTAAACCGT 238
QY 41 ThrLeuAlaValProProTTPileGluTyrGlnHisHisLysProProPheThrAsnLeu 60
DB 239 ACCTTGCTGCTCCCTTCCTTGGATTGAGTACCAAGCTCAGCCCTCCAGGCTTACCATCGGGTCATC 298
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
DB 299 CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 358
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTTPProGluLysArgVal 100
DB 359 AGCTTGGAGGATTCATGGAGAAGCTGGCACCCACCACTGGGCCCTCGAGAAGCGGGTG 418
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
DB 419 GCATCTGCTTGGAGTGGAGCCAGCCAGATAGAGAAGCTGCCCATGAG 478
QY 121 GluGlyAsnProPheGlyProPheTTPAspGlnPheHisValSerPheAsnLysSerGlu 140
DB 479 GAAGGAAACCCCTTTGGGCCATTCCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538
QY 141 LeuPheThrGlyLeSerPheSerAlaSerTyrArgGluGlnTTPSerGlnArg 158
DB 539 CTTTTTACAGGCAATTCCTTCAGTGTCTCTACAGAGAACAATGGAGCCAGAG 592

RESULT 10

US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18161
; LENGTH: 1209

```
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18161

Alignment Scores:
Pred. No.: 9,42e-94 Length: 1209
Score: 831.00 Matches: 167
Percent Similarity: 61.23% Conservative: 62
Best Local Similarity: 44.65% Mismatches: 119
Query Match: 41.99% Indels: 26
DB: 26

US-09-774-954-2 (1-365) x US-11-097-143-18161 (1-1209)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetClyArgPheGlyAenGlnAlaAasp 26
DB 82 GATCCCAATGGCTACCTCAGTCTACTGCGGTATGGGACGCTTGGCAACACCGGCGAC 141

QY 27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
DB 142 CACTTCTGGGATCATTTGGCTTCGCCAAGCGCTTAATCGCACCTGTATCTCGCGCGG 201

QY 47 TrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHisValSerTyrGlnLys 66
DB 202 TGGGTGGAGTATCGT-----AGGGGTGAACTCGATCCCGTCAGGTACCGTTCAACACA 255

QY 67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAaspPheMet 86
DB 256 TATTTTGAAGTGGAGCCCTCGAAGGAATACCATCGCGTCATCCATGGGAGATTTTCATG 315

QY 87 GluLysLeuAlaProThrHisTyrProGluLysArgValAlaTyrCysPheGlu--- 105
DB 316 TGGCACCTGGCGACGACATTTGGCCAGAAATCGGAGCGAGTGTCAITTTGCTACAGGAA 375

QY 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
DB 376 CGATATAGCCTTCAGCAGGAGAGAACGATCCAGACAAGCCCAATTCGCCACGCCAAGAT 435

QY 122 GlyAsnProPheGlyProPheThrAspGlnPheHisValSerPheAsnLysSerGluLeu 141
DB 436 GGCAATCCTTTTGGTCCCTTTTGGGACACATTTTTCACATAGACTTTGTGCGGTGAGAGTTC 495

QY 142 PheThrGlyLysSerPheSerAlaSerTyrArgGluGln-----TyrSerGlnArg 158
DB 496 TATGCGGCACCTTCATTTTGATGTGCATCATAGCAACGAGCTGCCAAGTGGCAGACCAA 555

QY 159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
DB 556 TATCCTGCAGAAATCATATCCCGTACTCGGTTCACCGGAGCTCGGCTAGTTTTCCTGTT 615

QY 179 LeuGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr 198
DB 616 CAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATATGGGAGCA 675

QY 199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
DB 676 TCTAAGGATTTTCATCCGAGAGCAGTTCCTCGGGGTGCGCTTTTGGGCATTCATCTCGC 735

QY 218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237
DB 736 AACGGTATCGATTGGGTGAGNGCTGTGAGCACGCTCAAGGAT-----AGCCAGCAT 786

QY 238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
DB 787 CTGTTTGCCTCGCGCAGTGTCTGGGTATATAAATAAAGCTGGTGCA---CTCTACCCG 843

QY 258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
DB 844 GAGCTCTGATGCCCTCAAGAGGGCGGATCATCGCCAGCTTAAAGAGAACCATTTAAGAAC 903

QY 278 Leu-----AspAlaGlnSerValTyrValAlaThrAspSerGlu 290
DB 904 GTGCGCCAAACTCAGCGGACCAAGAAATCAATCAGTTTTCGTGGGTGCGTCAAGCTCCAA 963

QY 291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysValValSerLeu 309
DB 964 CACATGATTGGTGAACATAACACGGCCCTTAGTCGATGGGCATGATGTGCACAGCTG 1023

QY 310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
DB 1024 CCGGAGGATGATCCTTACCTGGACCTTGGCCATTTCTCGGACAGTCGAACCACTTTATCGGC 1083

QY 330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
DB 1084 AACTGTATATCTCTTACTCGGCATTCGAAAAAAGGAAAGAGATGTGCACGGTTTCCA 1143

QY 350 SerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
DB 1144 TCGTACTTCTGGGATTC-----CCCAAGAAAAAGGAT 1176

RESULT 11
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18160
; LENGTH: 3264
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.: 1.09e-88 Length: 3264
Score: 796.00 Matches: 167
Percent Similarity: 58.27% Conservative: 62
Best Local Similarity: 42.49% Mismatches: 119
Query Match: 40.22% Indels: 45
DB: 26 Gaps: 10

US-09-774-954-2 (1-365) x US-11-097-143-18160 (1-3264)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
DB 2183 GATCCCAATGGCTACCTCAGTCTACTGCTCGGTATGGTAAAGTTTCACTTGGAAAGCAACC 2124

QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisP 28
DB 2123 GAAAAACGACAACTAAACCGCGCTTTTCAAGGACGCTTTGGCAACCGCCGACCACTT 2064

QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProTrpIle 48
```



```
Db 2063 CTGGGATCAATGGCCCTTGGCAAGCGCTTAATCGCACCCCTGATCGCGCGCGTGGT 2004
Qy 48 eGluTyGlnHisHisLysProProPheThrAsnLeuHisValSerTyGlnLysTyPh 68
Db 2003 GGAGTATCGT-----AGGGGTGAATCGCATCCCGTCAGGTACCGTTCAACACATATT 1950
Qy 68 eLysLeuGluProLeuGlnAlaTyHisArgValIleSerLeuLysAspPheMetGlu 88
Db 1949 TGAATGGAGCCCTTGAAGAAATACATCGCTCATCATCCATGCGAGATTTTCATGTGGCA 1890
Qy 88 sLeuAlaProThrHisTrpProGluLysArgValAlaTyCysPheGlu----- 105
Db 1889 CTGGCCGACGACATTTGGCCAGAAATCGGAGCGAGTGCATTTTGTCTACAGGAAGATA 1830
Qy 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123
Db 1829 TAGCCTTCAGCAGGAGAAAGACGATCCAGACAGCCCAATTGCCACGCCAAGATGCCAA 1770
Qy 123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
Db 1769 TCTTTTGGTCCCTTTGGACACTTTTCATAGACTTTTGGCGTCAGAGTCTCTATGC 1710
Qy 143 rGlyIleSerPheSerAlaSerTyArgGluGln-----TrpSerGlnArgPheSe 160
Db 1709 GCCACTTCAATTTGTATGTGCATCATAGCAAGCGTCCCAAGTGGCAGACCAATATCC 1650
Qy 160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuG 180
Db 1649 TGCAGAAATCATATCCGTACTCCGGTTCACCGGAGCTCCGGCTAGTCTTCTGCTCAGCT 1590
Qy 180 uGluHisArgProLeuGlnLysTyMetValTrpSerAspGluMetValLysThrGlyG 200
Db 1589 AGAGACTGCAGCTGCAGCGCTACTTTCAGTGGAGTCAACGGTATAGGAAGCATCTAA 1530
Qy 200 uAlaGlnIleHisAlaHisLeuValArg---ProTyValGlyIleHisLeuArgIleG 219
Db 1529 GGATTTCAATCCGAGCAGAGTGTCTCCGGGTGCTTTTGGGCATTCATCTGCGCAACGG 1470
Qy 219 ySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMe 239
Db 1469 TATCGATTGGGTGAGAGCTGTGAGCAGCTCAAGGAT-----AGCCAGCATCTGTT 1419
Qy 239 tAlaSerProGlnCysValGlyTySerArgSerThrAlaAlaProLeuThrMetThrMe 259
Db 1418 TGCTTCGCGCAGTGTCTGGGCTATAAAATGAACGTGGTGCA---CTCTACCGCGAGCT 1362
Qy 259 tCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeu-- 278
Db 1361 CTGATCCCTCCAGGAGCGATCATCCCGCAGCTAAAGAGAACCATTAAGAAGCTGGC 1302
Qy 279 -----AspAlaGlnSerValTyValAlaThrAspSerGluSerTy 292
Db 1301 CCAAACTCAGCCGCAACGAATAATCAATCAGTTTTCGTGGCGTCAGACTCCAATCAT 1242
Qy 292 rValProGluLeuGln---GlnLeuPheLysGlyLysValLysValSerLeuLysPr 311
Db 1241 GATTGGTGAATAAACACGCCCTTAGTCGATGGGCGATCAGTGTGCAACAGCTCGCGGA 1182
Qy 311 oGluValAlaGlnValAspLeuTyIleLeuGlyGlnAlaAspHisPheIleGlyAsnCy 331
Db 1181 GGATGATCCTTACCTGATCTGGCATCTCTCGACAGTGAACCACTTATCGCAACTG 1122
Qy 331 sValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgProSerSe 351
Db 1121 TATATCTCTTACTCGGCATTCGAAAGGAAAGGAAACGAGATGTGCACGGTTTTCATCGTA 1062
Qy 351 rPhePheGlyMetAspArgProProLysLeuArgAsp 363
Db 1061 CTTCTGGGGATTC-----CCCAAGGAAAGGAT 1034
```

RESULT 12

US-11-097-143-18082/c

```
/ Sequence 18082, Application US/11097143
/ Publication No. US20050208558A1
/ GENERAL INFORMATION:
/ APPLICANT: Venter, J. Craig
/ APPLICANT: et al.
/ TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
/ TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
/ TITLE OF INVENTION: DETROPHILA GENES.
/ FILE REFERENCE: CL000728
/ CURRENT APPLICATION NUMBER: US/11/097,143
/ PRIOR FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: 60/157,832
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: 60/160,191
/ PRIOR FILING DATE: 1999-10-19
/ PRIOR APPLICATION NUMBER: 60/161,932
/ PRIOR FILING DATE: 1999-10-28
/ PRIOR APPLICATION NUMBER: 60/164,769
/ PRIOR FILING DATE: 1999-11-12
/ PRIOR APPLICATION NUMBER: 60/173,383
/ PRIOR FILING DATE: 1999-12-28
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 18082
/ LENGTH: 3793
/ TYPE: DNA
/ ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.: 4,49e-52 Length: 3793
Score: 506.50 Matches: 103
Percent Similarity: 56.38% Conservative: 34
Best Local Similarity: 42.39% Mismatches: 59
Query Match: 25.59% Indels: 47
DB: 26 Gaps: 5

US-09-774-954-2 (1-365) x US-11-097-143-18082 (1-3793)
Qy 7 AspProAlaGlyTyIleLeuTyCysProCysMet----- 18
Db 675 GATCCCAATGGCTACCTCACCTACTGTCCGTATCGGTAACTTTTCATTGGAAGCAACC 616
Qy 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db 615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCGAGCGCCACTT 556
Qy 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProTrpI 48
Db 555 CTGGGATCATTCGGCTTCGCCAAGCGCTTAATCGCACCTGATCCTGCCCGCGTGGT 496
Qy 48 eGluTyGlnHisHisLysProProPheThrAsnLeuHisValSerTyGlnLysTyPh 68
Db 495 GGAGTATCGT-----AGGGGTGAATCGCATCCCGTCAGGTACCGTTCAACACATATT 442
Qy 68 eLysLeuGluProLeuGlnAlaTyHisArgValIleSerLeuLysAspPheMetGlu 88
Db 441 TGAAGTGCAGCCCTCAAGGAATACCATCGCTCATCCATGCGCAGATTTTCATGTGGCA 382
Qy 88 sLeuAlaProThrHisTrpProGluLysArgValAlaTyCysPheGlu----- 105
Db 381 CTGGCCGACGACATTTGGCCAGAAATCGGAGCGAGTGTCAATTTTGTCTACAGGAACGATA 322
Qy 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123
Db 321 TAGCCTTCAGCAGGAGAAAGACGATCCAGACAGCCCAATTGCCACGCCAAGATGGCAA 262
```



```
QY 123 nProPheGlyProPheTyrAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
Db 261 TCCTTTTGGTCCCTTTTGGGACACTTTTACATAGACTTTGTGGCTCAGAGTTCTATGC 202
QY 143 rGlyIleSerPheSerAlaSerTyrArgGluGln-----TyrSerGlnArgPheSe 160
Db 201 GCCACTTCATTGATGTGTCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAATATCC 142
QY 160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
Db 141 TGCAGAATCATATCCCGTACTCCGCTTACCGGAGCTCCGGCTAGTTTCTCTGTCAGCT 82
QY 180 uGluHisArgProLeuGlnLysTyrMetValTyrSer-AspGluMetValLysThrGlyG 200
Db 81 AGAGAACTGCAACTGACGCGCTACTTGCAGTGGAGTCAACGGT----- 38
QY 200 luAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlyS 220
Db 37 -----ATAGGGAAGCATCTAAGGATTTCAT 13
QY 220 erAsp 221
Db 12 CCGAG 8
RESULT 13
US-09-764-891-6032
; Sequence 6032, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6032
; LENGTH: 10331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6032
Alignment Scores:
Pred. No.: 5,16e-41 Length: 10331
Score: 424.00 Matches: 83
Percent Similarity: 98.81% Conservative: 0
Best Local Similarity: 98.81% Mismatches: 1
Query Match: 21.42% Indels: 0
DB: 10 Gaps: 0
US-09-774-954-2 (1-365) x US-09-764-891-6032 (1-10331)
QY 223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
Db 2491 AAGAACCCCTGTGCATCTGTGAAGACGGAGCTCGAGCTCGCACTTCATGGCTCTCCG 2550
QY 243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
Db 2551 CAGTGTGTGGCTACAGCGCGCAGCACAGCGGCCCTCCAGATGACTATGTGCTGCTCC 2610
QY 263 AspLeuLysGluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAlaGlnSer 282
Db 2611 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGCTGGATGCCAGTCG 2670
QY 283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPheLysGly 302
Db 2671 GTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTGCCAAGCTCTTCAAAGGG 2730
QY 303 LysValLysVal 306
Db 2731 AAGGTATGTGTG 2742
```

```
RESULT 14
US-09-918-995-19437
; Sequence 19437, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19437
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19437
Alignment Scores:
Pred. No.: 4.16e-39 Length: 479
Score: 393.00 Matches: 76
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.86% Indels: 0
DB: 10 Gaps: 0
US-09-774-954-2 (1-365) x US-09-918-995-19437 (1-479)
QY 290 GluSerTyrValProGluLeuGlnLeuPheLysGlyLysValLysValSerLeu 309
Db 31 GAGAGTTATGTGCTGAGCTCCACAGCTTTCAAGGGAAGGTGAAGTGTGTGAGCTG 90
QY 310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
Db 91 AAGCCTGAGGTGGCCAGGTGCGACCTGTACATCTCGGCCAAGCCGACCACTTTATTGGC 150
QY 330 AsnCysValSerSerPheThrAlaPheValLysArgGluAspLeuGlnGlyValGPro 349
Db 151 AACTGTGTCTCTCTCTCACTGCTTTGTGAAGGGAGCGGACCTCCAGGGAGGCCCG 210
QY 350 SerSerPheGlyMetAspArgProProLysLeuArgAspGluPhe 365
Db 211 TCCTCTTTCTTCGCGCATGACAGAGCCCTTAAGCTGCGGACCGATTTC 258
RESULT 15
US-11-097-143-18169/c
; Sequence 18169, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18169
; LENGTH: 3567
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18169

Alignment Scores:
Pred. No.: 1.14e-37 Length: 3567
Score: 392.00 Matches: 88
Percent Similarity: 57.80% Conservative: 38
Best Local Similarity: 40.37% Mismatches: 76
Query Match: 19.81% Indels: 16
DB: 26 Gaps: 6

US-09-774-954-2 (1-365) x US-11-097-143-18169 (1-3567)

QY 155 TrpSerGlnArgPheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAla 174
DB 3567 TGGCAGACCAATATCTCGAGAAATCATATCCGCTACTCGGTTTCACCGAGCTCCGGCT 3508
QY 175 GlnPheProValLeuGluHisArgProLeuGlnLysTyrMetValTyrSerAspGlu 194
DB 3507 AGTTTCTCTGTTTCAGCTAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGG 3448
QY 195 MetValLysThrGlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGly 213
DB 3447 TATAGGAGAGCATCTAAGAGTTTCATCCGAGACAGTTCCTCGGGTGCCTTTTGGGC 3388
QY 214 IleHisLeuArgIleGlySerAspTyrLysAsnAlaCysAlaMetLeuLysAspGlyThr 233
DB 3387 ATTCACTCGCAACGGTATCGATGGTGGTGGAGCGCTGGAGCAGCTCAAGGAT----- 3334
QY 234 AlaGlySerHisPheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAla 253
DB 3333 ---AGCCAGCATCTGTTTGCCTCGCCGCTGCTTGGCTATAAAATGAACGTTGTGCA 3277
QY 254 ProLeuThrMetThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeu 273
DB 3276 ---CTCTACCCGGAGCTCTGCATGCCCTCCAGGAGGCGATCATCCGCCAGCTAAAGAGA 3220
QY 274 TrpValArgSerLeu-----AspAlaGlnSerValTyrValAla 286
DB 3219 ACCATTAGAACGTGGGCAAACTCAGCCGGCAACAGCAAAATCAATCAATTTTCGTGGCG 3160
QY 287 ThrAspSerGluSerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLys 305
DB 3159 TCAGACTCCAATCACATGATTGGTGAACATAACACCGCCCTTAGTCGATGGGATCAGT 3100
QY 306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
DB 3099 GTGCACAAGCTGCCGGAGGATGATCTTACCTGGACTTGGCCATTTCTCGACAGTCCGAAC 3040
QY 326 HisPheIleGlyAsnCysValSerPheThrAlaPheValLysArgGluArgAspLeu 345
DB 3039 CACTTATCGCACTGATATATCTTCTTACTCGCATTCGAAATGAGGAAACGAGATGTG 2980
QY 346 GlnGlyArgProSerSerPhePheCysMetAspArgProProLysLeuArgAsp 363
DB 2979 CACGGTTTTCCTGCTACTTCTGGGGATTC-----CCCAAGAAAAGGAT 2935

```

Search completed: October 26, 2005, 15:59:38
Job time : 1594.35 secs

DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:59:38 ; Search time 6375.13 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPMGR.....OGRPSFFGMDRPPKLRDSEF 365

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool.h/US09774954/runat 25102005.105432.6406/app_query.fasta_1.917
-DB=EST -QPMT=FASTAP -SUFFIX=ret -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09774954 @CGN 1.1 5334 @runat 25102005.105432.6406 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_ges1.*
9: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	1555	3	CR617950 full-length
2	1832	92.6	3611	3	AK081059 Mus muscu
3	1829	92.4	2616	3	AK044629 Mus muscu
4	1525	77.1	1141	5	BX353274 BX353274
5	1394	70.4	1091	5	BM909055 AGENCOURT
6	1356	68.5	801	5	BI144264 BI144264
7	1355	68.5	947	4	BI410783 BI410783
8	1350.5	68.2	889	4	BI524031 BI524031
9	1339	67.7	988	5	BQ964179 AGENCOURT

10	1329	67.2	841	5	BX345989 BX345989
11	1316	66.5	915	5	BQ936634 AGENCOURT
12	1304	65.9	898	5	BX449172 BX449172
13	1304	65.9	942	7	CO774833 ILLUMIGEN
14	1272	64.3	719	1	AL135434 DKFZP7620
15	1237	62.5	808	4	BG747760 B02705255
16	1196	60.4	874	4	BI757540 B03029693
17	1178	59.5	921	1	AJ454995 AJ454995
18	1176	59.4	928	7	CF239348 AGENCOURT
19	1167	59.0	919	5	BU155142 AGENCOURT
20	1164.5	58.8	809	4	BG473023 B02514976
21	1164	58.8	927	6	CA974668 AGENCOURT
22	1152	58.2	705	7	CN348925 170003328
23	1126	56.9	753	4	BG424731 B02453575
24	1118	56.5	760	5	BU055535 UI-M-FOO-
25	1099	55.5	775	5	BU485326 B03843502
26	1087	54.9	689	4	BI559635 B03252494
27	1073	54.2	763	7	CK366990 AGENCOURT
28	1044	52.8	819	7	CF219216 AGENCOURT
29	1027	51.9	590	4	BG894786 B55525 MA
30	1021	51.6	666	5	BP459416 BP459416
31	1019	51.5	736	5	BP457865 BP457865
32	1001	50.6	561	4	BM798261 K-EST0081
33	987	49.9	644	7	CN459072 UI-M-RNO-
34	976	49.3	731	1	AJ453431 AJ453431
35	976	49.3	4147	3	AK053648 Mus muscu
36	964	48.7	695	2	BB617910 BB617910
37	963	48.7	664	6	BY728337 BY728337
38	961	48.6	743	5	BP686346 BP686346
39	957	48.4	624	7	CN348922 170004249
40	952	48.1	641	2	BB611933 BB611933
41	944	47.7	661	2	BB642654 BB642654
42	943	47.7	699	1	AU137536 AU137536
43	941	47.5	662	4	BM426089 pgf2n.pk0
44	940	47.5	746	5	BP699469 BP699469
45	936	47.3	545	7	CN348928 170005313

ALIGNMENTS

RESULT 1
CR617950
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK

CR617950 1555 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DJ008Yf12 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).

CR617950 GI:50498757
CR617950.1 GI:50498757
HTC; CNSLT_CDNA.
Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1555)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished

Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
Genoscope.
2 (bases 1 to 1555)

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source

Location/Qualifiers
1..1555
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D008YF12"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="PCWSPORT_6"

ORIGIN

Alignment Scores:
Pred. No.: 3,53e-204 Length: 1555
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-2 (1-365) x CR617950 (1-1555)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 109 ATGCTTGGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCGCCCTGCATGGGGCGC 168
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 169 TTGGGACCCAGCCGCGATCACTCTCTGGGCTCTCTGGCACTTTGGCAAGCTGCTAAACCGT 228
QY 41 ThrLeuAlaValProProTrpIleGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 229 ACCTTGCTGCTCTCTCTGAGTGGATACGATCATCAAGCCCTCTTTCACCACTC 288
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGlnProLeuGlnAlaTyrHisArgValile 80
Db 289 CATGTGCTTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 348
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTTPProGluLysArgVal 100
Db 349 AGCTTGGAGATTTTCATGGAGAGCTGGCACCCACCTGCGCCCTCGAGAGCGGGTG 408
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 409 GCATACTGCTTTGAGGTGGCAGCCAGAGCCAGATGAAGAAGACGTGCCCATGAAG 468
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 469 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTCAATGAGTTTCAACAAGTCGGAG 528
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 529 CTTTTTACAGCATTTCTTCAAGTCTTCTTACAGAGAACAAATGGAGCCAGAGATTTCT 588
QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db 589 CCAAGGAACATCGGTGTTTCCCTGCGAGAGCCAGCCAGCCAGTTCCTCGCTAGAG 648
QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 649 GAACACAGGCCACTACAGAGTACATGTATGGTCAGACGAAATGTGAGACGGGAGAG 708
QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 709 GCCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 768
QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 769 GACTCGAAGAACGCTGTGCCATGTCTGAAGGACGGGACTGCAGGCTCGCATCTTCATGCC 828
QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 829 TCTCCGAGTGTGTGGGTACAGCCGACAGCAGCGGCCCTCCACGATGATGATGTGC 888
QY 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db 889 CTGCTTGACCTGAGAGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 948

QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 949 CAGTCGGTCTACGTTGCTACATTCGAGAGTTAIGTCTGAGCTTCCACAGCTCTTC 1008
QY 301 LysGlyLysValLysValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 1009 AAAGGAAGGTGAAGGTGGTGAAGCTTGAAGTGGCCAGGTCGACCTGTACATC 1068
QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 1069 CTGGCCAAAGCCGACCACTTTATTTGGCACTGTGTCTCTCTTCTGCTTGTGAAG 1128
QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
Db 1129 CGGAGGGGGGACCTCCAGGGAGGCGCTCTTCTTCTTCTGCGCATGACAGGCCCTTAAG 1188
QY 361 LeuArgAspGluPhe 365
Db 1189 CTGGGAGCAAGTTTC 1203
RESULT 2
AK081059 3611 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
DEFINITION enriched library, clone:B9J0076G19 product:protein
O-fucosyltransferase 1, full insert sequence.
ACCESSION AK081059
VERSION AK081059.1 GI:26099647
KEYWORDS HTC, CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, I., Harada, N.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12000000
6 (bases 1 to 3611)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
92279253
10349636

2
Carninci, P., Shibata, Y., Hayateu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

20495374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

6 (bases 1 to 2616)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Segabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Retina RNA was provided by Dr. Stefano Gustincich (Department of
Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA
02115, USA) whose assistance is gratefully acknowledged. Please
visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1.. 2616
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"

/db_xref="FANTOM_DB:A930028F21"
/db_xref="taxon:10090"
/clone="A930028F21"
/tissue_type="retina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
34..1215
/note="unnamed protein product; protein
O-fucosyltransferase 1 (MGDI|GI:2153207, GB|NM_080463,
evidence: BLASTN, 99%, match=1290)
putative"
/codon_start=1
/protein_id="BAC32009.1"
/db_xref="GI:26336653"
/translation="MGAAWAPPHLLLRASFLILLLLPLRGSAGSWDLGALLYLPQ
CMGRGNQADHFLGSLAFALNRLTAVPWEYOHKHPPTNLHVSYQKFKLEPLQ
AYHRVUSLDPENLAPSWPEKRVAYCFVAAQSRSDKTKPMKEGNPFGPMDQF
HVSFNKSLPTGISFSAYKEQWQRPFAKHPVLALPGAPQFVLEHRELQKVMV
WDEMRIRGEALISSHLVRYGVIHLRIGSDWKACAMLDGTAGSHFASPOCVGYS
RSTAPLTNCLPDLKLEIQRAVTLWVRALNARSVYIATDSSESVSRIQOLFQKVRV
VSLKPEVAQIDLYILGQDRHFGNCVSSFTAVKREDLHGRQSSFGMDRPSQLRDE
F"

CDS

Alignment Scores:
Pred. No.: 1.79e-187 Length: 2616
Score: 1829.00 Matches: 335
Percent Similarity: 96.14% Conservative: 14
Best Local Similarity: 92.29% Mismatches: 14
Query Match: 92.42% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-2 (1-365) x AK044629 (1-2616)

QY 3 AlaglySerTrpAspProAlaGlyTrpLeuLeuTy-CysProCysMetGlyArgPheGly 22
DB 124 GCGGGCTCTGGGACCTGGCGGTACTGCTCTACTGCTCTGCTGGCGGTGGG 183
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaGlySerLeuAsnArgThrLeu 42
DB 184 AACCGGCTGATCATCTCTGGGCTCTGCTGGCATTTGGGAAGCTCTCAACCTCCTTG 243
QY 43 AlaValProTrpIleGluTrpGlnHisGlyProPhePheThrAsnLeuHisVal 62
DB 244 GCTTACTCTCATGGATTGATACCAATCAACATCAAGCTCTTCAACCACTCCATGTG 303
QY 63 SerTyrglnIleTyrrPheIleLeuGluProLeuGlnAlaTyrrHisArgValIleSerLeu 82
DB 304 TCCTACCAAAAGTACTTCAAACTGGAGCTCTCCAGGCTTACCATCGGTTGTGAGCTG 363
QY 83 GluAspPheMetGluIleLeuAlaProThrHisTrpProGluIleArgValAlaTyrr 102
DB 364 GAGGACTTTCATGAAAATCTGGCACTCTCCCTGGCGCTTGAAGCGAGTGGCATAC 423
QY 103 CysPheGluValAlaAlaGlnArgSerProAspIleIleTyrrCysProMetIleGluGly 122
DB 424 TGTCTTGGGTGGCAGCCAGCAAGTCTCTGATGAAGACATGTCTCCATGAAGGAAGA 483
QY 123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnIleSerGluLeuPhe 142
DB 484 AATCTTTTGGCCATCTCTGGACCCAGTTTCATGTGATTTCAATAAGTCAGAACTGTT 543
QY 143 ThrGlyIleSerPheSerAlaSerTyrrArgGluGlnTrpSerGlnArgPheSerProIle 162
DB 544 ACAGGCATTTCTTTCAGCGCTCTCTCAAAAGAACATGGAGCCAGAGATTTCTCTCAAAA 603
QY 163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluHis 182
DB 604 GAGCATCTCTGCTCGCCTGCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 663
QY 183 ArgProLeuGlnIleTyrrMetValTrpSerAspGluMetValIleTyrrGluAlaGln 202
DB 664 AGGGAGCTCCCAAGATGATGTTGTGGTGTGATGATGATGATGATGATGATGATGATG 723

/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores: 6.47e-155 Length: 1141
Pred. No.: 1525.00 Matches: 311
Score: 87.19% Conservativeness: 2
Percent Similarity: 86.63% Mismatches: 40
Best Local Similarity: 77.06% Indels: 10
Query Match: 5 Gaps: 4
DB:

US-09-774-954-2 (1-365) x BX353274 (1-1141)

```

Qy 203 ILeHsAlaHsLeuValArgProTyrValGlyLeHsLeuArgIleGlySerAspTtp 222
Db 724 ATTCAGTTCCCACTCGTCGCGCCCTATGTGGGATTCATCTGCATGCTCCGACTGG 783
Qy 223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
Db 784 AAGAAATGCTGTGCCATGCTGAAGATGGAATGCAGGGTCACTTCATGGCTTCCCT 843
Qy 243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
Db 844 CAGTGTGTGGGCTATAGCCGACGACACGACCCCTCTACCATGACCATGTGCTCCCT 903
Qy 263 AspLeuLysGluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAlaGlnSer 282
Db 904 GACCTGAAGGAATCCAGCGGGCTGTGACCGCTTTGGGTGAGAGACATGAATGCAGATCG 963
Qy 283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPheLysGly 302
Db 964 GTCTACATGCCACAGACTCTGAGAGCTACGTGTGAGAGATCCAGCAGCTCTTCAAGAC 1023
Qy 303 LysValLysValValSerLeuLysProGluValAlaGlnValAlaAspLeuTyrIleLeuGly 322
Db 1024 AAGGTAGGGTGTGTAGCTGAAACCCGAGGTGGCCAGATCGACCTGTATCTCTCGGC 1083
Qy 323 GlnAlaAspHisPheIleGlyAsnCysValSerPheThrAlaPheValLysArgGlu 342
Db 1084 CAGCTGACCACTTCATTTGAAATGTGTCTCTCTGTTTCACTGCTTCTGTTGAGAGGGAG 1143
Qy 343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
Db 1144 CGGACCTTCATGGAGGAGTGTCTCTTCTTGGCATGACAGACCTCCCGAGTTCTGG 1203
Qy 363 AspGluPhe 365
Db 1204 GATGAATTT 1212

RESULT 4
BX353274 1141 bp mRNA linear EST 08-APR-2004
LOCUS BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC003VJ11 5-PRIME, mRNA sequence.
ACCESSION BX353274
VERSION BX353274.2 GI:46290116
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1141)
Li, W.B., Gruber, C., Jesses, J. and Polaynes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30371746.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
9980.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DC003CE06QPlac=9980.f.
FEATURES
Location/Qualifiers
1..1141
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC003VJ11"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"

```

source

```
QY 280 aGInSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPh 300
Db 925 CCAGTCGGCTACGTTGCTACTGATCCGAGAGTTATGTTCTAGAGTCCAAACAGCTCTT 984
QY 300 eLysGlyLysValValSerLeuLysProGluValAlaGlnValAspLeuTyr11 320
Db 985 CAAAGGAGAGGTGAAGGTGGTGAG-CTKAGCVCYTAGGK-GSCCAGGTGCACTGT---AM 1039
QY 320 eLeuGlyGlnAlaAaphisPheIleGlyAenCysValSerPheThrAlaPheVally 340
Db 1040 ATCTCGGCACCGACACTTATTTGGCAATKT-KTTTCTCYTAT---TGCTTTGTAA 1095
QY 340 sArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArgPro 358
Db 1096 G---GGGAGGAGACMYCAGGGG-----RCGGTTTTTTTTTGGMATGAAGGCC 1141

RESULT 5
BM909055
LOCUS
DEFINITION BM909055 1091 bp mRNA linear EST 12-MAR-2002
5', mRNA sequence.
AGENCOURT_6611915 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431669
ACCESSION BM909055
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1091)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1904 row: e column: 14
High quality sequence stop: 629.
FEATURES
Location/Qualifiers
1..1091
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5431669"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_98"
/note="Organ: brain; Vector: pOTB7; Site: 1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 1,146-140 Length: 1091
Score: 1394.00 Matches: 290
Percent Similarity: 79.78% Conservative: 6
Best Local Similarity: 78.17% Mismatches: 25
Query Match: 70.44% Indels: 50
DB: 5 Gaps: 5
US-09-774-954-2 (1-365) x BM909055 (1-1091)
QY 1 MetProAlaGlySerTyrAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
```

```

77 ATGCCCGCGGGCTCTCGGACCCGCGGTTACTGCTACTGCTCCCTCGCATG----- 130
21 PheGlyAsnGlnAlaAaphisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
130 ----- 130
41 ThrLeuAlaValProProTrrP1leGluTyrGlnHisLysProProPheThrAsnLeu 60
131 -----GCT 133
61 -HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgVal11 80
134 CCATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCAT 193
80 eSerLeuGluAspPheMetGluLysLeuAlaProThrHisTrrProProGluLysArgVa 100
194 CAGCTTGGAGGATTTTCATGAGAAGCTGGCACCCACCCACCTGGCCCCCTGAGAAGCGGT 253
100 lAlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLy 120
254 GGCATATCTGCTTGGAGGTGGCAGCCAGCAGATAGAGATAGAGAGCGTCCCATGAA 313
120 sGluGlyAsnProPheGlyProPheTrrPaspGlnPheHisValSerPheAsnLysSerGl 140
314 GGAAGGAAACCCCTTTGGCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGA 373
140 uLeuPheThrGlyLleSerPheSerAlaSerTyrArgGluGlnTrrSerGlnArgPheSe 160
374 GCTTTTTCAGGCAATTTCTTTCAGTGCTTCTTACAGAGAAACAATGGAGCCAGAGATTTTC 433
160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
434 TCCAAAGGAAACATCGGTGCTTGGCTTCCAGAGAGCCAGCCAGCTTCCCGTCTAGA 493
180 uGluHisArgProLeuGlnLysTyrMetValTrrSerAspGluMetValLysThrGlyGl 200
494 GGAACACAGCCACATACAGAGATGATGTTGTCAGACAAATGGTGAAGACGGAGA 553
200 uAlaGlnLleHisAlaHisLeuValArgProTrrValGlyLleHisLeuArglleGlySe 220
554 GGCCAGATTCATCCACCTGTCGGCCCTATGTTGGGCAATTCATCTGCCCATTTGGCTC 613
220 rAspTrrLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAl 240
614 TGACTGGAAGAACGCTCTGTCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 673
240 aSerProGlnCysValGlyTrrSerArgSerThrAlaAlaProLeuThrMetThrMetCy 260
674 CTCTCCGCACTGTGTGGGTACAGCCGACAGCAGCGGCGCCCTCCAGATGATATGTG 733
260 sLeuProAspLeuLysGluLleGlnArgAlaValLysLeuTrrValArgSerLeuAspAl 280
734 CTGCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCTGGATGC 793
280 aGlnSerValTrrValAlaThrAspSerGluSerTrrValProGluLeuGlnLeuPh 300
794 CCAGTCGGCTACGTTGCTACTGATTCGAGAGATTATGCTCGAGCTCCACAGCTCTT 853
300 e--LysGlyLysValLysValVal-SerLeuLysProGlu-ValAlaGlnValAspLeuT 319
854 TAAAGGGAAGGGGGAAGGTGTGTGAACCTGAGGGGCCCGCCCTGAGGCTCCACCTGT 913
319 YrrleLeuGlyGlnAlaAaphisPhe-IleGlyAsnCysValSerSerPheThrAla--- 337
914 AAATTTCTTCGCAAGCCGAACCTTTAAATTTGGGAATGTGGCCCTCTTCCCGGGGCC 973
338 PheValLysArgGluArgAsp---LeuGlnGlyArg-----ProSerSerPhePhe 353
974 TTTGAAAAAGGAAACGGGACCCCTCTCAGGGGAGGGGGCGGCTTTTCTCTCT-----TTC 1027
354 GlyMet-AspArgProPro 359
||| :|||:|||||
||| :|||:|||||
```



```

Db      1028 GCAATGGAAAGCCCCC 1046
RESULT 6
BP144264
LOCUS   BP144264 full-length enriched swine cDNA library, adult ovary Sus
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa (pig)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS  Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
          Okumura,N., Hamasima,N. and Awata,T.
TITLE    PRDE (Pig EST Data Explorer): construction of a database for ESTs
          derived from porcine full-length cDNA libraries
JOURNAL  Nucleic Acids Res. 32 (1), D484-D488 (2004)
COMMENT  Contact: Hirohide Uenishi
          Animal Genome Laboratory, Genome Research Department
          National Institute of Agrobiological Sciences
          2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
          Tel: +81-29-838-8627
          Fax: +81-29-838-8627
          Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.
FEATURES
source
location/Qualifiers
1..801
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="OVR010034G09"
/tissue_type="ovary"
/dev_stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult
ovary"
ORIGIN
Alignment Scores:
Pred. No.: 9.66e-137 Length: 801
Score: 1356.00 Matches: 246
Percent Similarity: 97.37% Conservative: 13
Best Local Similarity: 92.48% Mismatches: 7
Query Match: 68.52% Indels: 0
DB: 5 Gaps: 0
US-09-774-954-2 (1-365) x BP144264 (1-801)
Qy      66 LysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPhe 85
Db      3 AAGTACTTCAAGCTGGAGGCCCTCCAGGTCTACACCGGGGTATCAGCTGGAGACTTC 62
Qy      86 MetGluLysLeuAlaProThrHisTyrProProGluLysArgValAlaTyrCysPheGlu 105
Db      63 ATGGAGAAGCTGGCACCACCACTGGCCCCCTGAGAGCGGGTGGCATACTGTTTGGAG 122
Qy      106 ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluLysAspProPhe 125
Db      123 GTGGCAGCACAGCGAAGCCAGATAAGAGACATGCCCCCATGAAGGAGGAACCCCTTT 182
Qy      126 GlyProPheTyrAspGlnPheHisValSerPheAsnLysSerGluLeuPheThrGlyIle 145
Db      183 GGCCCATTTTGGGATCAGTTTCAATGTCAGTTTCAACAGTCAGACTTTTTCAGGAGCAT 242
Qy      146 SerPheSerAlaSerTyrArgGluGlnTyrSerGlnArgPheSerProLysGluHisPro 165
Db      243 TCCTTCAGTGCCTCTCTACAAAGACCAAGTGGATCCAAAGATTTTCTCAAAGGAACATCCA 302
Qy      166 ValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHisArgProLeu 185
Db      303 GTGCTTGGCCTCCAGAGGCGCGGCTCAGTTCTCCCGTCTCTGGAGGAACATAGGCACCT 362
Qy      186 GlnLysTyrMetValTyrSerAspGluMetValLysThrGlyGluAlaGlnIleHisAla 205
Db      363 CAGAAATACATGGTGGTGCAGAGAGATGGTGAGACAGGGGAGGCCAGATCCGTGCC 422
Qy      206 HisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTyrLysAsnAla 225
Db      423 CACCTCATCCGGCCCTACGTGGGCATTACCTGCGCATTTGGCTCTGACTGGAAGACGCG 482
Qy      226 CysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerProGlnCysVal 245
Db      483 TGGCCCATCTGAAGACGCGCACTGGCGGCGCTCACTTATGGCCTCCCGCAGTGGGTG 542
Qy      246 GlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuProAspLeuLys 265
Db      543 GGTTCAGCGGCCACACCGCGCGCCGCTCACCATGACCATGTGCCTCCCGACCTGAAG 602
Qy      266 GluIleGlnArgAlaValLysLeuTyrValArgSerLeuAspAlaGlnSerValTyrVal 285
Db      603 GAGATCAGCGCGCCCTCTGAAGCTCTGGGTGACGGCGCTGAGCGGCCAGTCCGTCTATC 662
Qy      286 AlaThrAspSerGlnSerTyrValProGluLeuGlnGlnLeuPheLysGlyLysValLys 305
Db      663 GCCACGAGCTCTGAGAGTTACCTGCCGAGATCCAGCAGCTCTTCAAAGGAAAGGTGAAG 722
Qy      306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
Db      723 GTGGTGAAGTCTGAACCTGAGGTGGCCAGATGACCTGTATCTCTTGGCCAGGCTGAC 782
Qy      326 HisPheIleGlyAsnCys 331
Db      783 CACTTTATTGGCAACTGT 800
RESULT 7
BP140783
LOCUS   BP140783 947 bp mRNA linear EST 14-AUG-2001
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-x@mail.nih.gov
          Tissue Procurement: Gilbert Smith, Ph.D.
          cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
          Bonaldo, Ph.D.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LIA011290 row: 0 column: 13
          High quality sequence start: 21
          High quality sequence stop: 919.
FEATURES
source
location/Qualifiers
1..947
/organism="Mus musculus"
/mol_type="mRNA"

```


QY 203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTTP 222
 Db 752 ATCAGTGGCCACCTCGTCGGCCCTATGTGGCAATTCATCTGGCATTGGCTCCGACTGG 811
 QY 223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
 Db 812 AAGATGCTGTGGCCATCTGGAGATGGAATCGCAGGGTCACACTTCATGGCTTCCCT 871
 QY 243 Gln-CysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPr 262
 Db 872 TCNNITGTGGGCTATAGCCGACGACAGCAGCCCTCTCTAC-ATGACCATGTGCTCC 930
 QY 262 oaspleuLysGluIle 267
 Db 931 TGACCTGAAGGAATC 946
 RESULT 10
 EX345989
 LOCUS
 DEFINITION BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) linear EST 23-APR-2004
 Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
 ACCESSION
 VERSION EX345989.2 GI:46549776
 KEYWORDS EST.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 841)
 Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30377055.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 9980.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QP1&c=9980.f.
 FEATURES
 source
 1..841
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ008YF12"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_lines="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
 ORIGIN
 Alignment Scores:
 Pred. No.: 9,146-134 Length: 841
 Score: 1329.00 Matches: 242
 Percent Similarity: 99.18% Conservative: 0
 Best Local Similarity: 99.18% Mismatches: 2
 Query Match: 67.16% Indels: 1
 DB: 5 Gaps: 0
 US-09-774-954-2 (1-365) x BX345989 (1-841)
 QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20

109 ATCCCTGCGGGCTCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGCGC 168
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 Db 169 TTTGGGAACCCAGCCGATCACTTCTTTGGGCTCTCTGGCATTTGCCAAAGCTGCTAAACCGT 228
 QY 41 ThrLeuAlaValProProTTPIleGluTyrGlnHisHisLysPheProPheThrAsnLeu 60
 Db 229 ACCTTGGCTGTCTCTCTCTGGATTGAGTACCAAGCTTCAAGCCCTCTCTTTCACCAACCTC 288
 QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
 Db 289 CATGTGTCTTACCAAGAGTACTTCAAGCTGGAGCCCTCCAGGGCTTACCATCGGTCATC 348
 QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTTPProProGluLysArgVal 100
 Db 349 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCCTGAGCCCTGAGAAAGCGGGTG 408
 QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 Db 409 GCATACTGCTTTGAGGTGGCAGCCAGCGAAC-CCAGATAAGAGAGCGTCCCATGAAG 467
 QY 121 GluGlyAsnProPheGlyProPheTTPAspGlnPheHisValSerPheAsnLysSerGlu 140
 Db 468 GAAGGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAAGTCGGAG 527
 QY 141 LeuPheThrGlyLysPheSerAlaSerTyrArgGluGlnTTPSerGlnArgPheSer 160
 Db 528 CTTTTCACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACAATGGAGCCAGAGATTTTCT 587
 QY 161 ProLysGluHisProValLeuAlaLeuProGlyValProAlaGlnPheProValLeuGlu 180
 Db 588 CCAAGGAACAATCCGGTCTTGGCCCTGCCAGGAGCCCAAGCCAGTTCCTCCCTCTAGAG 647
 QY 181 GluHisArgProLeuGlnLysTyrMetValTTPSerAspGluMetValLysThrGlyGlu 200
 Db 648 GAACACAGGCCACTACAGAGTACATGTTGTCAGACGAAATGTTGAAGACGGGAGAG 707
 QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 Db 708 GSCCATGATTCATGCCCACCTTGTCCGGCCCTATGTGGCATTTTCATCTGCGCATTTGGCTCT 767
 QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 Db 768 GACTTGAAGAAACGCTGTGCCATCTGAAGACGGGACTGCAGGCTCGCACTTCATGGCC 827
 QY 241 SerProGlnCys 244
 Db 828 TCTCCGCAAGTGT 839
 RESULT 11
 EX345989
 LOCUS
 DEFINITION AGENCOURT_10032510 NCI CGAP Mam2 Mus musculus cDNA clone
 IMAGE:6485896 5', mRNA sequence.
 ACCESSION
 VERSION BQ936634
 KEYWORDS BQ936634.1 GI:22352017
 SOURCE EST.
 ORGANISM Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 915)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM14028 row: g column: 17
 High quality sequence stop: 698.

```

FEATURES
source
Location/Qualifiers
1. 915
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6485896"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	2,736-132
Score:	1316.00
Percent Similarity:	96.43%
Best Local Similarity:	93.65%
Query Match:	66.50%
DB:	5
Length:	915
Matches:	236
Conservative:	7
Mismatches:	9
Indels:	0
Gaps:	0

US-09-774-954-2 (1-365) x B0936634 (1-915)

Qy	3	AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly	22
Db	156	GCGGGCTCTCGGACCTGC CGGGTTACTCTGTCTACTCTCCCTGCATCGGGCGGCTTGGG	215
Qy	23	AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu	42
Db	216	AAC CAGCGCTGATCACATTCTTGGGCTCCCTGGCATTTTGGCAAGCTGTGAAACCGCACCTTG	275
Qy	43	AlaValProProTrpIleGluTyrGlnHisLysLysProPheThrAsnLeuHisVal	62
Db	276	GCTGTACCTCCATGGATTGAATACCAACATCACAAAGCCTCCTTTTCCAAACCTCCATGTG	335
Qy	63	SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValLleSerLeu	82
Db	336	TCCTACCAAAAGTACTTCAAACCTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCCGCGCTG	395
Qy	83	GluAspPheMetGluLysLeuAlaProThrHisTyrProGluLysArgValAlaTyr	102
Db	396	GAGGACTTCATGAAAATCTGGACCCCTCCCACTGGCCCCCTGAGAGCGAGTGGGCATAC	455
Qy	103	CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly	122
Db	456	TGCTTTGAGGTGGCAGCCAGCGAAGTCTCGATAAGAAAGACATGTCCCATGAAGGAAGGA	515
Qy	123	AsnProPheGlyProPheThrAspGlnPheHisValSerPheAsnLysSerGluLeuPhe	142
Db	516	AATCCTTTTGGGGCATTCCTGGGACCACTTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC	575
Qy	143	ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys	162
Db	576	ACAGCCATTTTCCCTTCAGCGCTCCTACAAAGAAACAAATGAGCCAGAGATTTCTCTGCAAAA	635
Qy	163	GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis	182
Db	636	GAGCATCTGTGCTCGCACTGCCTGGGGGCCAGACACATGTTCCCTGTCTCTGGAGAGAACAC	695
Qy	183	ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln	202
Db	696	AGGGAGCTCCAGAAAGTACATGTTGGTTCAGATGAGATGGTGAGGACGGGAGAGCCCTGT	755
Qy	203	IleHisAlaHisLeuValArgProTyrValGlyValIleHisLeuArgIleGlySerAspTrp	222

```
QY 41 ThrLeuAlaValProProTrrPilleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
|
|
|
Db 259 ACCTTGCGTGTCCCTCHTTGGATTGAGTACCAGCATCAAGAGCTCTCTTCCACCAACCTC 318
|
|
|
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
|
|
|
Db 319 CATGTGCTCTACAGAGTACTTAYAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 378
|
|
|
QY 81 SerLeuGluAapPheMetGluLysLeuAlaProThrHisLysTrpProGluLysArgVal 100
|
|
|
Db 379 AGCTTGGAGGATTCATGGAGAGCTGGCACCACCCACCTGCGCCCTTGAGAGCGGGTG 438
|
|
|
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
|
|
|
Db 439 GCATACCTGCTTTGAGGTGGCAGCCAGCGAAC -CCAGATAAGAAAGAGCTGCCCATGAAG 497
|
|
|
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
|
|
|
Db 498 GAAGGAAACCCCTTTGGCCCATCTCGGATCAGTTTCATGTAGTGTCAACAAGTCGGAG 557
|
|
|
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
|
|
|
Db 558 CTTTTTACAGGCATTTCTTCAGTGTCTGTACAGAGAACATGGAGCCAGAGATTTTCT 617
|
|
|
QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
|
|
|
Db 618 CCAAGGAACATCCCGTCTTGGCCCTGCCAGGAGGCCAGCCCAAGTTCCCGCTCTAGAG 677
|
|
|
QY 181 GluHisArgProLeuGlnLysTyrMetValTyrSerAspGlu-MetValLysThrGlyG1 200
|
|
|
Db 678 GAACACAGCCCACTACAGAGTACATGGTATGGTCAGACGAATGGTGAAGACGGGAGA 737
|
|
|
QY 200 uAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisArgIleGlySe 220
|
|
|
Db 738 CGCCACAGATTCAGCCCACTTGTCCGCGCCTATGTGGGCATTCATCTGCGCATTTGGCTC 797
|
|
|
QY 220 rAspTrrPlysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMet-A 240
|
|
|
Db 798 TGACTGGAAAGACGCGCTGTCCAATGCTGAGGACGGGACTCGAGGCTCGCACTTAATGGG 857
|
|
|
QY 240 laSerProGlnCysValGlyTyr-SerArgSerThrAlaAla 253
|
|
|
Db 858 CTYCTCCGAGTGTGTGGGCDACAGCGCCAGCACACSGGCC 898
|
|
|
RESULT 13
CO774833
LOCUS
DEFINITION
ILLUMIGEN MQ_51950 Katze_MNLV Macaca nemestrina cDNA clone
ISLUM:29960 5' similar to Bases 7 to 852 highly similar to human
POFUT1 (Hs.178292), mRNA sequence.
CO774833
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca nemestrina (pig-tailed macaque)
Macaca nemestrina
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
Cercopitheinae; Macaca.
1 (bases 1 to 942)
Katze,M.C., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
Large-scale Rhesus Macaque cDNA Sequencing
Unpublished (2003)
Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
```

```
PCR Primers
FORWARD: CCTCTACTAAAGGACACAAA
BACKWARD: CACTTAGGCGGAATTGGGTA
Insert Length: 942 Std Error: 0.00
Plate: CL000531 row: E column: 09
Seq primer: CCTCTACTAAAGGACACAAA
POLYA-No. Location/Qualifiers
1..942
/organism="Macaca nemestrina"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9545"
/clone="IBIUM:29960"
/sex="male"
/lab host="Electromax DH10B"
/clone_lib="Katze_MNLV"
/note="Organ: liver; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from CloneMiner cDNA Library
Construction kit (catalog #18249-029)"
```

FEATURES

source

ORIGIN

```
Alignment Scores: 5.8e-131 Length: 942
Pred. No.: 1304.00 Matches: 245
Score: 1304.00
Percent Similarity: 94.38% Conservative: 6
Best Local Similarity: 92.11% Mismatches: 11
Query Match: 65.89% Indels: 4
DB: 7 Gaps: 0

US-09-774-954-2 (1-365) x CO774833 (1-942)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|
|
|
Db 95 ATGCGCGCGGGTCTCTGGGACCCGCGCGTTACTGTCTACTGCGCCCTGCAATGGGCGC 154
|
|
|
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|
|
|
Db 155 TTTGGGAACACGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGCTCTTAACCGC 214
|
|
|
QY 41 ThrLeuAlaValProProTrrPilleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
|
|
|
Db 215 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCAAGCTCTCCCTTCCACCACTC 274
|
|
|
QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
|
|
|
Db 275 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCAATC 334
|
|
|
QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
|
|
|
Db 335 AGCTTGGAGATTTTATGGAGAGCTGGACCCACCCACTGCGCCCTTGAGAGCGGGTG 394
|
|
|
QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
|
|
|
Db 395 GCATACCTGCTTTGAGGTGGCAGCCCGAGAGCCAGATAAGAGACGCTGCCCATGAAG 454
|
|
|
QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
|
|
|
Db 455 GAAGGAAACCCCTTTGGCCCATCTCGGATCAGTTTCATGTAGTGTTCACACAGTCTGAG 514
|
|
|
QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
|
|
|
Db 515 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACATGGAGCCAGAGATTTTCT 574
|
|
|
QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
|
|
|
Db 575 CCAAGGAACATCCCGTCTTGGCCCTCGCGGAGCCCGCCAGTTCCTCTGTCTTGAA 634
|
|
|
QY 181 GluHisArgProLeuGlnLysTyrMetValTyrSerAspGluMetValLysThrGlyGlu 200
|
|
|
Db 635 GAACACAGACCACTGCAAAAGTACATGTGTATGTCAGACCAAAATGTCAGACCGGAGAA 694
|
|
|
QY 201 AlaGlnIleHisAlaHisLeuVal-ArgProTyrValGlyIleHisLeuArgIleGly-S 220
|
|
|
```

```

Db      695  GCCCAGAAATCCCGCCCACTTGTCGGGCGCTATGTGGGCAATTCCTCTACGTAATGGCTT 754
Qy      220  exAspTrpLysAenAlaCysAlaMeLeuLysAspGlyThrAlaGlySerHisPheMetA 240
Db      755  CTGACTGGAAATGCTGTGCTATGCTAAAGACGAACTGGCAGCTCGACTTCTCTGG 814
Qy      240  laSerProGlnCysValGly-TyrSerArgSerThrAlaAla-ProLeuThrMetThrMe 259
Db      815  CCTCTCCCAATGTGGGGCGCTACGCCCCCAACACAGCGCGCCCCCTCCGAGAACTAT 874
Qy      259  tCysLeuPro 262
Db      875  GTGCCTGCCT 884

RESULT 14
AL135434
LOCUS   DKFZp76200115 r1 762 (synonym: hm12) Homo sapiens cDNA clone
DEFINITION
ACCESSION   AL135434
VERSION     DKFZp76200115.5, mRNA sequence.
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
AUTHORS     This is the 5' sequence of the clone insert
TITLE       Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
JOURNAL     Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
COMMENT     sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No 81 sequence available.
            This clone (DKFZp76200115) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
            Location/Qualifiers
                1. 719
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="DKFZp76200115"
                    /tissue_type="melanoma (MeWo cell line)"
                    /dev_stage="adult"
                    /lab_host="DH10B"
                    /clone_lib="762 (synonym: hm12)"
                    /notes="Vector: pSport1; Site_1: NotI; Site_2: SalI"

FEATURES
    source
        1. 719
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp76200115"
            /tissue_type="melanoma (MeWo cell line)"
            /dev_stage="adult"
            /lab_host="DH10B"
            /clone_lib="762 (synonym: hm12)"
            /notes="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN
Alignment Scores:
Pred. No.:      1.16e-127      Length:      719
Score:          1272.00      Matches:      239
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    64.27%      Indels:      0
DB:             1      Gaps:      0

US-09-774-954-2 (1-365) x AL135434 (1-719)

Qy      121  GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db      2    GAAGGAAACCCCTTTGGGCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAGTCGGAG 61
Qy      141  LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160

```

```

Db      62  CTTTTTACAGGCATTTCTTCAGTGTCTCCTACAGAGAACAAATGGAGCCAGAGATTTTCT 121
Qy      161  ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db      122  CCAAGGAAACATCCGGTGTCTTGCCTGCCAGAGAGCCCGCCAGTTCCTCCGCTCTAGAG 181
Qy      181  GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db      182  GAACACAGGCCCACTACAGAAATGATGATGGTATGGTACAGCGAAATGGTGAAGACGGAGAG 241
Qy      201  AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db      242  GCCCAGATTCATGCCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGGCGATTTGGCTCT 301
Qy      221  AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db      302  GACTGGAGAAAGCCCTGTGCCATGCTGAAGACGGGACTGCGAGGCTCGCACTTCATGGGCC 361
Qy      241  SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db      362  TCTCCGACGTGTGGGCTACAGCGCGCAGCACAGCGGCCCGCCCTCAGCATGACTATGTGTC 421
Qy      261  LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db      422  CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 481
Qy      281  GlnSerValTyrValAlaThrAspSerGluSerTyrValProGlnLeuGlnLeuPhe 300
Db      482  CAGTCGGTCTACGTTCTACTGATCCGAGAGTTATGTGCTCTCTCTCTCTCTCTCTCTCT 541
Qy      301  LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db      542  NAAGGGAAGGTGAAGGTGGTGGAGCTTGAAGCTGAGTGGCCAGCTCGACCTGTATCATC 601
Qy      321  LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db      602  CTCGGCAAGCCGACCACTTTATTGGCACTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 661
Qy      341  ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProPro 359
Db      662  CGGGAGCGGAGACCTCCAGGGGAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 718

RESULT 15
BG747760      808 bp      mRNA      linear      EST 15-MAY-2001
LOCUS         602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
DEFINITION    mRNA sequence.
ACCESSION     BG747760
VERSION       BG747760.1 GI:14058413
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Mammalia; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 808)
AUTHORS       NIH-MGC http://mgs.nci.nih.gov/
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Ling Hong/Rubin Laboratory
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1675 row: n column: 14
              High quality sequence stop: 808.
              Location/Qualifiers
                  1. 808
                      /organism="Homo sapiens"
                      /mol_type="mRNA"

```


/db_xref="taxon:9606"
 /clone="IMAGE:4842061"
 /tissue type="normal pigmented retinal epithelium"
 /lab host="DH10B (phage-resistant)"
 /clone lib="NIH_MGC 43"
 /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library. |"

ORIGIN

Alignment Scores:

Pred. No.: 9.11e-124 Length: 808
 Score: 1237.00 Matches: 234
 Percent Similarity: 96.71% Conservative: 1
 Best Local Similarity: 96.30% Mismatches: 5
 Query Match: 62.51% Indels: 4
 DB: 4 Gaps: 0

US-09-774-954-2 (1-365) x BG747760 (1-808)

QY	1	MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	82	ATGCCCGCGGGCTCTGGACCCGGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC	141
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	142	TTTGGGACAGCGCGATCATCTTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	201
QY	41	ThrLeuAlaValProTrpTrpIleGlyTyrGlnHisLysLysProPheThrAsnLeu	60
DB	202	ACCTTGGCTGTCCTCTTGGATTGAGTACGACATCAAGCCCTCTTTACCAACCTC	261
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle	80
DB	262	CATGTGCTCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC	321
QY	81	SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal	100
DB	322	AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCATGCCCCCTGAGAAAGCGGGTG	381
QY	101	AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys	120
DB	382	GCATCTGCTTTGAGGTGGCAGCCAGCGAAGCCAGATAAGAAAGCGTCCCCCATGAAG	441
QY	121	GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu	140
DB	442	GAAGGAAACCCCTTTGGCCCATTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG	501
QY	141	LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer	160
DB	502	CTTTTACAGGCATTTCCTCAGTGTCTCTCAGAGAACATGGAGCCAGAGATTTTCT	561
QY	161	ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu	180
DB	562	CCAAAGAAATCCGGTGCTTGGCCCTCCAGGAGCCCGACGATTCCTCCCTTAGAG	621
QY	181	GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr-GlyI	200
DB	622	GAACACAGGCCCACTACAGAGTACATGGTATGTCAGACGAATGGTGAAGACGGGGAGA	681
QY	200	uAlaGlnIleHisAlaHisLeuValArgProTyrVal-GlyIleHisLeuArgIleGlyS	220
DB	682	GGCCAGATTTCATGCCCAACTGGTCCGACCTATGTTGGGCATCAATCTGCGCATTTGTC	741
QY	220	erAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySer-HisPheMet	239
DB	742	T-GACTGGAAGAAGCCCTGTGCCATCTGGAAGACGGGGACTGAGGCTCGGCATTTCATG	800

QY 240 Ala 240
 DB 801 GCT 803
 Search completed: October 26, 2005, 15:17:12
 Job time : 6398.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:43 ; Search time 45.7187 Seconds
(without alignments)
768.157 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: Pir1:.*
2: Pir2:.*
3: Pir3:.*
4: Pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	541.5	27.4	474	2	Ti5511
2	114	5.8	594	2	A86309
3	105.5	5.3	383	2	T14795
4	101.5	5.1	732	2	S07624
5	96.5	4.9	432	2	T37509
6	94.5	4.8	1209	2	T31857
7	93.5	4.7	458	2	G83690
8	91.5	4.6	1292	2	T31462
9	90.5	4.6	437	2	F59099
10	90	4.5	218	4	S37459
11	89.5	4.5	650	2	T38692
12	89	4.5	877	2	T03098
13	88.5	4.5	380	2	T40454
14	88	4.4	427	2	T04869
15	88	4.4	426	2	G64899
16	87.5	4.4	1246	2	D95211
17	87	4.4	407	2	F91080
18	87	4.4	520	2	G85925
19	87	4.4	492	2	T29439
20	86.5	4.4	1189	2	S00944
21	86.5	4.4	1246	2	G90897
22	86.5	4.4	1246	2	B95730
23	86.5	4.4	328	2	S39599
24	85.5	4.3	770	2	A12421
25	85.5	4.3	421	2	T47393
26	85	4.3	732	1	JC4655
27	85	4.3	859	2	S69700
28	85	4.3	410	2	E59168
29	84.5	4.3			

30	84.5	4.3	3519	2	S43048	polyketide synthas
31	84	4.2	1117	2	AE0075	hypothetical prote
32	83.5	4.2	283	2	S44848	K06H7.8 protein -
33	83.5	4.2	390	2	E86411	protein Flk23.18 [
34	83.5	4.2	518	2	JC7387	testis-abundant fi
35	83.5	4.2	1400	2	T33758	hypothetical prote
36	83	4.2	255	2	AH2479	transcription regu
37	83	4.2	349	1	E64096	UDPglucose-hexose-
38	83	4.2	554	2	S67452	hypothetical prote
39	83	4.2	803	2	AB2088	hypothetical prote
40	83	4.2	1242	2	AB0672	respiratory nitrat
41	82.5	4.2	263	2	JH0635	type II site-speci
42	82.5	4.2	534	2	F96713	unknown protein T6
43	82.5	4.2	659	2	AB1293	ABC transporter (p
44	82	4.1	520	2	T20226	hypothetical prote
45	82	4.1	623	2	B81037	lipopolysaccharide

ALIGNMENTS

RESULT 1

Ti5511
hypothetical protein C15C7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15511
R;Leimbach, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C15C7.
A;Reference number: Z18363
A;Accession: T15511
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-474 <UEI>
A;Cross-references: EMBL:U41528; NID:g1109795; PID:g1109800; PIDN:AAA83156.1; CESP:C15C7.
C;Genetics:
A;Gene: CESP:C15C7.1
A;Introns: 31/3; 67/2; 106/2; 166/1; 202/3; 235/3; 364/3; 410/3

Query Match	27.4%;	Score	541.5;	DB	2;	Length	474;
Best Local Similarity	36.2%;	Pred. No.	1.9e-38;				
Matches	125;	Conservative	60;	Mismatches	107;	Indels	53;
Gaps	11;						
QY	19	GREGNOADHPLSLAFKALNRLTAVPPMTEYQHHKPPFTNLHVSQYKFKLEPLQAYHR	78				
Db	166	GREGNOVDQGLGVIAFAKALDRTLVLPNFIEFKH---PETKM-IPFEFLQVGTVAKYTR	221				
QY	79	VISLEDFMEKLAPTHWPPEKRVAYCFEVAQAQRSPDKKTCPMKEGNGPPGPFWDQPHVSNFK	138				
Db	222	VVTMQEFTKKIMPTHFVGTPQA-IYDKSAEPG-----CHSKEGNGPPYWDQIDVSVFG	275				
QY	139	SEUFTGI--SFSASY---REQSQRSKPKHPVLALPGAAQPVLEHRLPKQYMWSD	193				
Db	276	DEYFGDIPGFDLQNGSRKKWLEKFPSEYPVLAFFSAPAPFPKSGKQVSIQKYLWSS	335				
QY	194	EMVKTGEAQIHAHLVPPYVGIHLRIGSDKNACAMLKDGTAGSHFMSAPQCVGVSRTAA	253				
Db	336	RITEQAQKIFISANLAKFPFAVHLNRNDWVRVCEHI--DTTNRPLFAEQCLGEGHILGT	394				
QY	254	PLTWTMCLPDLKEIQRAVKLVWRSLSAQSVYVATDSSEYVPELQQLFKGVKVVSLKPEV	313				
Db	395	-LYKEICSPSKQI-----LEQ-----IEAHRQBPDD	420				
QY	314	AQVDLYLQADHIFIGNCVSFTAFVVRERDLOG---RPSFFGM	355				
Db	421	MYTSLAIMGRADLFVGNVCVSTFSHIVKRERDHAGQSPRPSAFFGI	465				

RESULT 2

A86309
F20D23.3 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)

Thu Oct 27 11:17:19 2005

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A86309
R;Ethnology: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Rulzar, L.
Nature 408, 816-820, 2000
C;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.C.; Venter, J.C.; Davis, R.W.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86309
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-594 <STO>
A;Cross-references: UNIPROT:Q9SHI5; GB:AB005172; NID:G5734743; PIDN:AA05008.1; GSPDB:GN
C;Genetics:
A;Map position: 1
Query Match 5.8%; Score 114; DB 2; Length 594;
Best Local Similarity 20.1%; Pred. No. 0.1;
Matches 84; Conservative 50; Mismatches 160; Indels 124; Gaps 19;
QY 6 WDP-AGVLLYCPMGFGNQADHEL---GSLAFKLLNRTLAVP-PMIEYQHHKPPPTNL 60
Db 209 WKPRDKFLFAICLS--GOMSHLICKEHMFPAALLDRVILVIPSCKFDYQ----- 257
QY 61 HVSQYKFKLEPIQ---AHVRVLSLEDMKELAPTHWPEKRVAYCFEVAQSPDKKTC 117
Db 258 ---YDKVIDIERINTCLGRVTVISPDQKE-----IDKNNH-----DRFIC 298
QY, 118 PMKEGNFGFQWDPQHVSNKSELTGTSFSASVRQWQOR----- 158
Db 299 YVSSPQPC--YVDEHIIKLG--LGVSTGGKLEAPWSEDIKKTKTSQEVVEKPKSD 353
QY 159 -----FSPKEHFLVIALGPAQPFVLEERPLQKYMVSDEMVKTGBOAIIHAHLV 208
Db 354 DGVIAIGDVFADMEQDLVMPGGINHKCKTLTLEPSRLILVTAQRFIQ-----FLG 406
QY 209 RPYVGIHLR---IGSDKNACA---MLKDGTAGSHFMAQSPQ---CVYGRSTAAPLTM 257
Db 407 KNFISLHLRHGFLKFWLQALVAVMLDSVAHTQLIGIPEHLSTCNAKSPS----- 459
QY 258 TWCLPDLKEIQRVAVKLVRSLSAQSYYVATD--SSYVPELQQLFKGKVKVWSL---KPE- 312
Db 460 ---CFYPIQADCLSRMVERANAPVILSTDAESGTGLQSLVVVDGKVVPLVKRPQN 517
QY 313 -----VQVDLYILGQADHFICNGVSSFTAFVKEERDLQGRPS 350
Db 518 SAEKWDSLLVHNGIEDDSQVYAMLDTICAMSSVFIGASGTFTEDLRLKDWGTSS 575
RESULT 3
T14795
hypothetical protein DKFZp434E171.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14795
R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A;Reference number: Z18183
A;Accession: T14795
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-363 <POU>
A;Cross-references: EMBL:AL110285
A;Experimental source: adult testis; clone DKFZp434E171
C;Genetics:
A;Note: DKFZp434E171.1
Query Match 5.3%; Score 105.5; DB 2; Length 383;

Best Local Similarity 21.1%; Pred. No. 0.32;
Matches 87; Conservative 42; Mismatches 126; Indels 157; Gaps 18;
QY 3 AGSWDPAG-----YLLY--CPCMGFRGQADHFLGSLAFK 36
Db 16 AVSWPPASASGQEFWFGQSAADILSGAASRRRLLYDNNPPEG-FNLRRDYIRIASLLK 74
QY 37 LLNRT-----LAVPPMIEYQH-HKPPFTNLHVSQYKFKLEPLQAYHVRVLSLEDMKELAP 91
Db 75 TLLKTEWVLVLPMPGRGLYHQSPDIHQVRIPWSEFFDLFSLNKNIPVIEYEQF---IAE 131
QY 92 THWPPERVAYCFEVAQAQSPDKKTCMKEGN-----PFGPFWDDQFHVSNKSELTFTG-- 144
Db 132 SGGPFIDQV-YVLQSVAB-----GKSGTWEKVDERPDCIDQLLYSQDKHEYVRGWF 182
QY 145 -----ISFSASY-----REQWSQFS 160
Db 193 WGYEETRGVNSCLSVQGSASIVAPLLRLNTSARSVMDRAENLLHDHGGKEYWDTRRS 242
QY 161 PKEHPVIALPGAPAPFPVLEHR-----PLOCKYMWSDVMKVTGEAQIIHAHL 207
Db 243 -----WFAHLREVGDFRSHLNSDADRIIPQF--DWMKMKVKGLSA-----L 287
QY 208 VRPVGHLRIGSDWKNACMLKDGTAGSHFMAQSPQCVGYSRSTAAPLTMTCPLDLKEI 267
Db 288 GGPYLGVHLR-----RKDFIWHGR-----QDVPSLEGA 315
QY 268 QRAVKLVRSLSAQSYYVATDSYVPELQQLFKGKVKVSVLSKPEVAQVDLY 319
Db 316 VRKIRSLMKTHRLDKVFVATDAVR--KEYEELKLLPEMVRPEPTWEDELEY 365
RESULT 4
S07624
acylaminoacyl-peptidase (EC 3.4.19.1) - rat
N;Alternate names: acyl-peptide hydrolase; acylamino-acid-releasing enzyme
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 09-Jul-2004
C;Accession: A33706; S07624
R;Kobayashi, K.; Lin, L.W.; Yeaton, J.B.; Klickestein, L.B.; Smith, J.A.
J. Biol. Chem. 264, 8892-8899, 1989
A;Title: Cloning and sequence analysis of a rat liver cDNA encoding acyl-peptide hydrolase
A;Reference number: A33706; MUID:89255359; PMID:2722805
A;Accession: A33706
A;Molecule type: mRNA
A;Residues: 1-732 <KOB>
A;Cross-references: UNIPROT:P13676; GB:J04733; NID:G202931; PIDN:AAA88506.1; PID:G202932
A;Note: parts of this sequence, including the amino end of the mature protein, were deter
R;Lin, L.W.; Lee, F.J.S.; Smith, J.A.
Nucleic Acids Res. 17, 4397-4400, 1989
A;Title: Structural organization of the rat acyl-peptide hydrolase gene.
A;Reference number: S07624; MUID:89296508; PMID:2578023
A;Accession: S07624
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-732 <LIN>
A;Cross-references: EMBL:X14915
C;Genetics: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 386/2
C;Intron: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 386/2
C;Superfamily: acylaminoacyl-peptidase
C;Keywords: blocked amino end; homotetramer; hydrolase; omega peptidase
F;Modified site: blocked amino end (Met) (probably acetylated) #status experimental
F;18,291,443/Modified site: lysine derivative (Lys) #status experimental
F:587,675,707/Active site: Ser, Asp, His #status predicted
Query Match 5.1%; Score 101.5; DB 1; Length 732;
Best Local Similarity 22.9%; Pred. No. 1.6; Indels 139; Gaps 23;
Matches 81; Conservative 32; Mismatches 101;
QY 30 GSLAFAKLLN-----RTLAVPPMIEYQHHPFTNLHVSQYKFKLEPLQAYH-- 77
Db 405 GSAGSKLLTIDKDLVMAQFSTPLSPKLVGLFPFGKESQSVSWSLAEAFIPGIHWG 464

RESULT 11
S37459
self-incompatibility glycoprotein, mutant non-functional allele - Peruvian tomato (fragm
C;Species: Lycopersicon peruvianum (Peruvian tomato)
C;Date: 06-Jan-1995 #sequence_revision 31-Jan-1997 #text_change 09-Jul-2004
C;Accession: S37459
R;Royo, J.; Kunz, C.; Koyama, Y.; Anderson, M.; Clarke, A.E.; Newbiggin, E.
Proc. Natl. Acad. Sci. U.S.A. 91, 6511-6514, 1994
A;Title: Loss of a histidine residue at the active site of S-locus ribonuclease is assoc
A;Reference number: A58473; MUID:942294411; PMID:8022814
A;Accession: S37459
A;Molecule type: mRNA
A;Residues: 1-218 <ROY>
A;Cross-references: UNIPROT:O40245; EMBL:Z26581; NID:9404333; PIDN:CAA81332.1; PID:g4043
A;Experimental source: strain LA 2157, style tissue
A;Note: submitted to the EMBL Data Library, September 1993
A;Note: part of this sequence, including the amino end of the mature protein, was determ
A;Note: the source is designated as Solanum peruvianum
C;Keywords: angiosperm reproduction; glycoprotein
F;1-17/Domain: signal sequence (fragment) #status predicted <SIG>
F;18-218/Product: self-incompatibility glycoprotein #status predicted <MAT>
F;46;56/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.5%; Score 89.5; DB 4; Length 218;
Best Local Similarity 19.4%; Pred. No. 3.6;
Matches 45; Conservative 43; Mismatches 103; Indels 41; Gaps 10;
QY 62 VSYQKVFPLEQYHVRVSLDFMEKLAHPHWPPEKRVAYCFEVAQAQRSPDK 121
DB 4 VFMFLPALSPYGFLEL-----VSTWP-----ATYCYAGCSRR-----PIPN 45
QY 122 GNFGPPWDFHVSFNKSELFTGISASVREQWSQRFSPKEHPVLALPGAPQFVL-- 179
DB 46 NFTINGLWPD-----NKSVINLNCNF-AKKEDRYTKITDPKKSSEL-----DKRWPOLRY 94
QY 180 EHRPLQKYMVWSDVMVKTGEAQTIAHLVRYVGIHLRIGSDWKNAACMLKDGTAGSHFM 239
DB 95 EKLXGIEKQDLWEKFLKHGSCSINRYKQAYFDLAKMI---KDRFDLL--GTLRNGI 148
QY 240 ASPQCVGYRSTRAPLMTWCLPDLKIEIQRAVKLVWRSLSDAQSVYVATDS 291
DB 149 IPGSTVELDDIERAVKTVSIEVPSLKIQP-----LGNVELNEIGICLDPEA 196

RESULT 12
T38692
probable serine/threonine-specific protein kinase (EC 2.7.1.1-) SPAC3C7.06c [similarity]
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2004
C;Accession: T38692
R;Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21805
A;Accession: T38692
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-650 <BRO>
A;Cross-references: UNIPROT:O14132; EMBL:Z99568; PIDN:CAB16737.1; GSPDB:GN00066; SPDB:SP
A;Experimental source: strain 972h; cosmid c3C7
C;Genetics:
A;Gene: SPAC3C7.06C
A;Map position: 1
C;Superfamily: protein kinase homology
C;Keywords: phosphotransferase

Query Match 4.5%; Score 89; DB 2; Length 650;
Best Local Similarity 22.3%; Pred. No. 16;
Matches 71; Conservative 50; Mismatches 126; Indels 72; Gaps 16;
QY 4 GSWDPAGYL-----LYCPCMG--RFGNQADHFL--GSLAFKLLNRTLAVPPWIEYQHKK 54
DB 273 GIWDRAELLANKLGISLPKMAPLDFG---DLFSPFWNLAFASMLSQLKWDP-----AK 323

QY 55 PPFTNLHVSQYKFKLEPLQAYHVRVSLDFMEKLAHPHWPPEKRVAYCFEVAQAQRSPDK 114
DB 324 RPTAALCLDLE-----FCRVAPADAVASKEEVNKNDFRVSI5YFFPSSIPDE 373
QY 115 KTCPMKEGNPFGPFWDQFHVSNKSLFTGIGSFSASYREQWSQ- 427
DB 374 --CNTTEESRIINPSTSKFLKQKNG--FNG--FTKPKRSKRSKQKQKQKSSVATQFSEE 427
QY 163 EHPVLALPGAPAQFPVLEBHRP-----LOKYMVWSDVMVKTGEAQTIAHLVRYVYVGIH 215
DB 428 SEDIADISITSSITFFVPLPQIRPSTPLNLKRLNFISSSESDSTSPKAK---EFDRLPLPSTE 484
QY 216 LRIGSDWKNAACMLKDGTAGSHFMASPOCVGYRSTRAAPLMTWCLPDLKIEIQRAVKLVW 275
DB 485 FLVAIN-KSQEALLNN-----SPN-----SKSGSTQLSASTCLSLISQLSILSHE 530
QY 276 RSLDAQSVYVATDSYVVP 294
DB 531 DKRENQS--VNSESSKYSYP 547

RESULT 13
T03098
p97 protein - Toxoplasma gondii
C;Species: Toxoplasma gondii
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03098
R;Matsuura, T.; Kasper, L.
Mol. Biochem. Parasitol. 90, 403-413, 1997
A;Title: Molecular analysis and characterization of a protein involved in the replicator
A;Reference number: Z14838; MUID:98135655; PMID:9476788
A;Accession: T03098
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-877 <MAT>
A;Cross-references: UNIPROT:O15644; EMBL:AF005059; NID:G2581824; PIDN:AAC47857.1; PID:G25
A;Experimental source: strain RH
C;Function:
A;Description: involved in replication of intracellular Toxoplasma gondii
C;Superfamily: Toxoplasma gondii p97 protein

Query Match 4.5%; Score 88.5; DB 2; Length 877;
Best Local Similarity 19.9%; Pred. No. 27;
Matches 74; Conservative 44; Mismatches 119; Indels 135; Gaps 20;
QY 7 DPAGYLLYPCMGFGNQADHFLGSLAFKLLNRTLAVPPWIEYQH-HKPPF----- 57
DB 392 DPEGWKLKLAHMQRGRR--FPGD--FDTWLRIHQEAKEWRRVHSGPLNKLADPE 447
QY 58 -----TNLHVSQYKFKLEPLQAYHVRIS-----LEDPMKLAFT 92
DB 448 GWELFRLRSARGNTTVPATYAEWMKLPEQOVFVAVDAAWAAWRQTHPGPLKDLVLAQDAGT 507
QY 93 HWP-----PEKRVAYC-----FEVAQAQRSPDKKTCPMKE--GNPPGPFWDQ 131
DB 508 GWLYRAMLRADLDLDPDDLATWLDQSGKHILKAWQAQRDKHKGPLRRVLADPLG--W-- 563
QY 132 FHVSNKSELFTGISFSASYREQWSQRFSPKE-----HPVLALP----- 170
DB 564 --VLFKNMVESGASIEQTF--EWAQSKAPSGNEWINWRAHKGPLHEVLQDPGWLKLYR 620
QY 171 -----GAPAQF-----PVLEHHRPLQKYMVWSDVMVKTGEAQTIAHLV----- 208
DB 621 ENLVRKNVDVPOSFEDFLOQPDQQAATSKSSW-----RRVHRGPLTVVLADK 671
QY 209 ---RPVYGIHLRIG-----SDWKNAACMLKDGTAGSHFMASPOCVGYRSTRAAPLTMT 258
DB 672 EGWQLYNNWMEKLGHKIKSSFEWSSSAQMHK--GRAAREWRA-----WRRVHKGPLTTV 724
QY 259 MCLPDLKIEIQRA 270
DB 725 LADPDGWDLYKA 736

RESULT 14
T04054
branched-chain-amino-acid transaminase (BC 2.6.1.42) precursor, mitochondrial - fission
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T04054; T40571
R:Lyne, M.; Rajandream, M.A.; Barrall, B.G.; Baker, S.; Mungall, K.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z21931
A:Accession: T04054
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <LYN>
A:Cross-references: UNIPROT:O14370; EMBL:AL034382; PIDN:CAA22277.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972h; cosmid c428
R:Wood, V.; Rajandream, M.A.; Barrall, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21937
A:Accession: T40571
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <WOO>
A:Cross-references: EMBL:AL096788; PIDN:CAB46675.1; GSPDB:GN00067; SPDB:SPBC582.12c
A:Experimental source: strain 972h-; cosmid c582
C:Genetics:
A:Gene: SPBC428.02c; SPBC582.12c
A:Map position: 2
A:Genome: nuclear
A:Superfamily: branched-chain-amino-acid transaminase BAT1
C:Keywords: aminotransferase; mitochondrion

Query Match 4.4%; Score 88; DB 2; Length 380;
Best Local Similarity 19.6%; Pred. No. 9; 9; Gaps 19;
Matches 67; Conservative 55; Mismatches 121; Indels 98; Gaps 19;
QY 52 HHKPPPTNLHVSYQYKFLQLPQLQYHVRVLSLEDFMEKLAPTHWPPEKRVAYCFEVAQAORS 111
DB 8 HGKPMWDSHHKVTNNVKELPDKWKLKGEFTDMLIMKNREKWS-----TPEIV 62
QY 112 POKKTCPMKEGNPFGFWDQHVHFNKSELFTGTSFSAISYREQWSQSPKHPVLALPG 171
DB 63 PFGKLC-----FHPASSVPHYGF---ECFEGMK---AFRDE-----KG 94
QY 172 APAQFPVLEHRPLQKYMVWSDVMVKTG-----EAQIHAHLVRPVY----- 212
DB 95 VPRLF-----RPIKN-----AERMLSTGTSLPSFDPDAEL-AETIRKFAHENRWVPDQ 143
QY 213 ---GIHLR---IGSDWKNACAMLDKGTAGSHFMAFSPQCVGYSRSTAAPLWTMCLPDLKE 266
DB 144 RGXSLVIRPTFTIGTD--EALGVHCDNAMYLVIASVPGPYTSSGPKA---VKLCCS---- 194
QY 267 IQRAVKLWVRSLDQAQSVVATDSYVPELQQLFKGKVKVSLKPEVAQVDLYLGOADH 326
DB 195 -ESSVRAMPGGTGHYKL--GGNYAPSVLPQKEAAKGYAQI-----LMLYGDYD 241
QY 327 F--IG--NCVSSFTAFVKRERDLQGRPSFFGMDRPPKLRD 363
DB 242 ITEVGTMNC---FTVWINKGEKIIITAPLDGMILPGVTRD 279

RESULT 15
T04869
transforming protein myb homolog F28A21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 16-Aug-2004
C:Accession: T04869
R:Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A:Accession: T04869
A:Molecule type: DNA
A:Residues: 1-427 <BEV>

A:Cross-references: UNIPROT:Q9S7L2; EMBL:AL035526
A:Experimental source: cultivar Columbia; BAC clone F28A21
C:Genetics:
A:Map position: 4
A:Introns: 226/2; 268/3
A:Note: F28A21.180
C:Superfamily: myb DNA-binding repeat homology
F:212-263/Domain: myb DNA-binding repeat homology <MYB>
Query Match 4.4%; Score 88; DB 2; Length 427;
Best Local Similarity 20.3%; Pred. No. 11;
Matches 71; Conservative 53; Mismatches 111; Indels 114; Gaps 19;
QY 56 PFTNLHVSYQK---YFKLEP-----LQA-----YHRVISLED----- 84
DB 84 PFE--HCSYQENWVDFYETKLNLMNHHHFOAVENSYFTRNHHHHEINLAVDEHDDPWDL 141
QY 85 ----FMEKLAPTHWPPEKRVAYCFEVAQAORSQSPDKKTCPMKEGNPFGFWDQFHYSENKSE 140
DB 142 QNNMMWRMIFPDYFPTT-----TFPMNFVMPDEISCVSADNDCYRA-----TSPNKT 191
QY 141 LF-----TGISFSAISYRE-----QWS-----ORFSPKE--HPVLALPGAP 173
DB 192 PFLTRKLSLSS 251
QY 174 AQFPVLEEH---RPLQKYMVWSDVMVKTGSAQIHAHLVRPVYGIHLRIGSDWKNACAMLK 230
DB 252 GKQCRERWNNHLRPDIKKTWSBE-----DRVLIPEHKEIGNKWAIEAKRLP 299
QY 231 DGTAGS---HFMAFSPQCVGYSRSTAAPLWTMCLPDLKEIQRAVKLWVRSLDQAQSVVAT 287
DB 300 GRTENSIGNHNATKRRQFQSKKCKRSKYPRPSLLQD-----YIKSLNMGALMA-- 347
QY 288 DSESYVP---ELQQLFKGKVKVSLKPEVAQVDLYLGOADHFGICNCV 332
DB 348 ---SSVPARGRRRESNKKDQVVAVEKKKEBEVY--QQ-DRIVPECV 390

Search completed: October 25, 2005, 15:36:41
Job time : 59.7187 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 59.9589 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-2

Perfect score: 1979

Sequence: 1 MPAGSWDPAGLYLPCPCMR.....QGRPSFFGMDRPPKLRDF 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*

2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*

3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*

4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*

5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	365	3	US-08-978-741-2
2	1979	100.0	365	3	US-09-333-729A-3
3	1979	100.0	397	3	US-08-978-741-6
4	1979	100.0	397	3	US-09-333-729A-7
5	1844	93.2	343	3	US-08-978-741-17
6	1844	93.2	343	3	US-09-333-729A-13
7	837	42.3	399	4	US-09-270-767-45921
8	541.5	27.4	474	3	US-08-978-741-8
9	541.5	27.4	474	3	US-09-333-729A-12
10	353	17.8	61	3	US-08-978-741-9
11	353	17.8	61	3	US-09-333-729A-4
12	343	17.3	200	4	US-09-270-767-61459
13	295	14.9	61	3	US-08-978-741-3
14	295	14.9	61	3	US-09-333-729A-5
15	101	5.1	28	3	US-08-978-741-14
16	101	5.1	28	3	US-09-333-729A-14
17	92.5	4.7	1271	4	US-09-540-236-3740
18	92.5	4.7	2074	4	US-09-491-356C-9
19	87.5	4.4	344	3	US-09-298-886-11
20	87.5	4.4	344	4	US-09-999-672-11
21	87	4.4	555	4	US-09-248-796A-20471
22	86.5	4.4	771	2	US-08-742-753-2
23	86.5	4.4	1189	1	US-08-602-737-2
24	86.5	4.4	1189	2	US-08-980-071-2
25	86.5	4.4	1189	2	US-08-980-071-4
26	86.5	4.4	1189	2	US-08-980-071-6
27	86.5	4.4	1189	2	US-08-980-071-8

28	86.5	4.4	1189	2	US-08-980-071-10	Sequence 10, Appl
29	86.5	4.4	1189	2	US-08-980-071-12	Sequence 12, Appl
30	86.5	4.4	1189	2	US-08-980-071-59	Sequence 59, Appl
31	86.5	4.4	1189	2	US-08-980-071-61	Sequence 61, Appl
32	86.5	4.4	1189	2	US-08-757-536-2	Sequence 2, Appl
33	86.5	4.4	1189	2	US-08-757-536-4	Sequence 4, Appl
34	86.5	4.4	1189	2	US-08-757-536-6	Sequence 6, Appl
35	86.5	4.4	1189	2	US-08-757-536-8	Sequence 8, Appl
36	86.5	4.4	1189	2	US-08-757-536-10	Sequence 10, Appl
37	86.5	4.4	1189	2	US-08-757-536-12	Sequence 12, Appl
38	86.5	4.4	1189	3	US-09-314-093-2	Sequence 2, Appl
39	86.5	4.4	1189	3	US-09-314-093-4	Sequence 4, Appl
40	86.5	4.4	1189	3	US-09-314-093-6	Sequence 6, Appl
41	86.5	4.4	1189	3	US-09-314-093-8	Sequence 8, Appl
42	86.5	4.4	1189	3	US-09-314-093-10	Sequence 10, Appl
43	86.5	4.4	1189	3	US-09-314-093-12	Sequence 12, Appl
44	86.5	4.4	1189	3	US-09-314-093-59	Sequence 59, Appl
45	86.5	4.4	1189	3	US-09-314-093-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1

US-08-978-741-2

; Sequence 2, Application US/08978741

; Patent No. 6100076

; GENERAL INFORMATION:

; APPLICANT: Yang Wang, Michael W. Spellman

; TITLE OF INVENTION: O-Fucosyltransferase

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/978,741

; FILING DATE: 26-No. 6100076-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/792498

; FILING DATE: 31

; ATTORNEY/AGENT INFORMATION:

; NAME: Svoboda, Craig G.

; REGISTRATION NUMBER: 39,044

; REFERENCE/DOCKET NUMBER: P1041P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650/225-1489

; TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 365 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

US-08-978-741-2

Query Match 100.0%; Score 1979; DB 3; Length 365;

Best Local Similarity 100.0%; Pred. No. 2.4e-214;

Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGLYLPCPCMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60

Db 1 MPAGSWDPAGLYLPCPCMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60

Qy 61 HVSQYKFKLEPLQAVHRVLSLEDFMEKLAHPHPPEKRVAYCFEVAQAQRSPDKKTCPMK 120

Db 61 HVSYQYFKLEPQAHYHRVISLEDFMEKLAHTHWPPEKRVAYCFEVAQAQSPDKKTCPMK 120
QY 121 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 180
Db 121 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 180
QY 181 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
Db 181 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
QY 241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 300
Db 241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 300
QY 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
Db 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
QY 361 LRDEF 365
Db 361 LRDEF 365

RESULT 2

US-09-333-729A-3

; Sequence 3, Application US/09333729A

; Patent No. 6270987

; GENERAL INFORMATION:

; APPLICANT: Wang, Yang

; TITLE OF INVENTION: O-Fucosyltransferase

; FILE REFERENCE: P1041PDI-Substitute

; CURRENT APPLICATION NUMBER: US/09/333,729A

; CURRENT FILING DATE: 1999-06-15

; PRIOR APPLICATION NUMBER: US 08/798,741

; PRIOR FILING DATE: 1997-11-26

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO 3

; LENGTH: 365

; TYPE: PRT

; ORGANISM: Homo Sapien

; US-09-333-729A-3

Query Match 100.0%; Score 1979; DB 3; Length 365;

Best Local Similarity 100.0%; Pred. No. 2.4e-214;

Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGRCFGNQADHFLGSLAFKILNRTLAVPPWIEYQHKKPPTNL 60
Db 1 MPAGSWDPAGYLLYPCMGRCFGNQADHFLGSLAFKILNRTLAVPPWIEYQHKKPPTNL 60
QY 61 HVSYQYFKLEPQAHYHRVISLEDFMEKLAHTHWPPEKRVAYCFEVAQAQSPDKKTCPMK 120
Db 61 HVSYQYFKLEPQAHYHRVISLEDFMEKLAHTHWPPEKRVAYCFEVAQAQSPDKKTCPMK 120
QY 121 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 180
Db 121 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 180
QY 181 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
Db 181 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
QY 241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 300
Db 241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 300
QY 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
Db 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
QY 361 LRDEF 365

Db 361 LRDEF 365
RESULT 3
US-08-978-741-6
; Sequence 6, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Mupatin (Genentech)
; CURRENT APPLICATION DATA: US/08/978,741
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-6
Query Match 100.0%; Score 1979; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.8e-214;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGRCFGNQADHFLGSLAFKILNRTLAVPPWIEYQHKKPPTNL 60
Db 33 MPAGSWDPAGYLLYPCMGRCFGNQADHFLGSLAFKILNRTLAVPPWIEYQHKKPPTNL 92
QY 61 HVSYQYFKLEPQAHYHRVISLEDFMEKLAHTHWPPEKRVAYCFEVAQAQSPDKKTCPMK 120
Db 93 HVSYQYFKLEPQAHYHRVISLEDFMEKLAHTHWPPEKRVAYCFEVAQAQSPDKKTCPMK 152
QY 121 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 180
Db 153 EGNPFQGFWDQPHVSNFKNSELTGTSFSASYREQMSQSPKHPVLALPGAPQAPFPVLE 212
QY 181 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 240
Db 213 EHRPLOKYMVWSDVMVKTGEAQIHAIHLVVRPYVGIHLRIGSDWKNAACMLKDGTAGSHFMA 272
QY 241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 300
Db 273 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRVAVKLWVRSIDAQSVVATDSYVPELQQLF 332
QY 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
Db 333 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 392

QY 361 LRDEF 365
 Db 393 LRDEF 397

RESULT 4
 US-09-333-729A-7
 ; Sequence 7, Application US/09333729A
 ; Patent No. 6270987
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Yang
 ; APPLICANT: Spellman, Michael W.
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; FILE REFERENCE: P1041P1D1-Substitute
 ; CURRENT APPLICATION NUMBER: US/09/333,729A
 ; CURRENT FILING DATE: 1999-06-15
 ; PRIOR APPLICATION NUMBER: US 08/798,741
 ; PRIOR FILING DATE: 1997-11-26
 ; NUMBER OF SEQ ID NOS: 21
 ; SEQ ID NO 7
 ; LENGTH: 397
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Plasmid insert encoded protein.

US-09-333-729A-7

Query Match 100.0%; Score 1979; DB 3; Length 397;
 Best Local Similarity 100.0%; Pred. No. 2.8e-214;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLYCPCMGFGNOADHFLGSLAFKLLNRTLAVPPWIEYQHKKPFTNL 60
 Db 33 MPAGSWDPAGYLYCPCMGFGNOADHFLGSLAFKLLNRTLAVPPWIEYQHKKPFTNL 92

QY 61 HVSYQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAQAQSPDKTCPMK 120
 Db 93 HVSYQYKFKLEPLQAYHRVLSLEDFMEKLAFTHPPEKRVAYCFEVAQAQSPDKTCPMK 152

QY 121 EGNPFGFWDQFHVSNKSLFTGTSFASYSREQWSQRFSPKHPVLALPGAPQFPVLE 180
 Db 153 EGNPFGFWDQFHVSNKSLFTGTSFASYSREQWSQRFSPKHPVLALPGAPQFPVLE 212

QY 181 EHRPQKYMWSDEMVKTGAEQIAHILVRPVYVGIHLRIGSDWKNACMLKDGTSAGHFWA 240
 Db 213 EHRPQKYMWSDEMVKTGAEQIAHILVRPVYVGIHLRIGSDWKNACMLKDGTSAGHFWA 272

QY 241 SPOCVGYSRSTAAPLTWTMCLPDLKEIQRAVKLWVRSLSDAQSVYVATDSYVPELQQLF 300
 Db 273 SPOCVGYSRSTAAPLTWTMCLPDLKEIQRAVKLWVRSLSDAQSVYVATDSYVPELQQLF 332

QY 301 KGKVKVSLKPEVAQVDLYILGQADHFGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
 Db 333 KGKVKVSLKPEVAQVDLYILGQADHFGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 392

QY 361 LRDEF 365
 Db 393 LRDEF 397

RESULT 5
 US-08-978-741-17
 ; Sequence 17, Application US/08978741
 ; Patent No. 6100076
 ; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California

COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WinPatIn (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,741
 FILING DATE: 26-No. 6100076-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/792498
 FILING DATE: 31
 ATTORNEY/AGENT INFORMATION:
 NAME: Svoboda, Craig G.
 REGISTRATION NUMBER: 39,044
 REFERENCE/DOCKET NUMBER: P1041P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-1489
 TELEFAX: 650/952-9881
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 343 amino acids
 TYPE: Amino Acid
 TOPOLOGY: Linear

US-08-978-741-17

Query Match 93.2%; Score 1844; DB 3; Length 343;
 Best Local Similarity 100.0%; Pred. No. 3.8e-199;
 Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 NOADHFLGSLAFKLLNRTLAVPPWIEYQHKKPFTNLHVSYQYKFKLEPLQAYHRVLSL 82
 Db 1 NOADHFLGSLAFKLLNRTLAVPPWIEYQHKKPFTNLHVSYQYKFKLEPLQAYHRVLSL 60

QY 83 EDFMEKLAFTHPPEKRVAYCFEVAQAQSPDKTCPMKEGNPFGFWDQFHVSNKSELF 142
 Db 61 EDFMEKLAFTHPPEKRVAYCFEVAQAQSPDKTCPMKEGNPFGFWDQFHVSNKSELF 120

QY 143 TGISFASYSREQWSQRFSPKHPVLALPGAPQFPVLEHRPQKYMWSDEMVKTGAEQ 202
 Db 121 TGISFASYSREQWSQRFSPKHPVLALPGAPQFPVLEHRPQKYMWSDEMVKTGAEQ 180

QY 203 IHAHLVRPVYVGIHLRIGSDWKNACMLKDGTSAGHFWASPOCVGYSRSTAAPLTWTMCLP 262
 Db 181 IHAHLVRPVYVGIHLRIGSDWKNACMLKDGTSAGHFWASPOCVGYSRSTAAPLTWTMCLP 240

QY 263 DLKEIQRAVKLWVRSLSDAQSVYVATDSYVPELQQLFKGKVKVSLKPEVAQVDLYILG 322
 Db 241 DLKEIQRAVKLWVRSLSDAQSVYVATDSYVPELQQLFKGKVKVSLKPEVAQVDLYILG 300

QY 323 QADHFGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
 Db 301 QADHFGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 343

RESULT 6
 US-09-333-729A-13
 ; Sequence 13, Application US/09333729A
 ; Patent No. 6270987
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Yang
 ; APPLICANT: Spellman, Michael W.
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; FILE REFERENCE: P1041P1D1-Substitute
 ; CURRENT APPLICATION NUMBER: US/09/333,729A
 ; CURRENT FILING DATE: 1999-06-15
 ; PRIOR APPLICATION NUMBER: US 08/798,741
 ; PRIOR FILING DATE: 1997-11-26
 ; NUMBER OF SEQ ID NOS: 21
 ; SEQ ID NO 13
 ; LENGTH: 343

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-333-729A-13

Query Match      93.2%; Score 1844; DB 3; Length 343;
Best Local Similarity 100.0%; Pred. No. 3,8e-199;
Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 NOADHFLGSLAFKALNRTLVAVPWIEYQHHKPPFTNLHVSYQYFKLEPLQAYHRVISL 82
DB 1 NOADHFLGSLAFKALNRTLVAVPWIEYQHHKPPFTNLHVSYQYFKLEPLQAYHRVISL 60
QY 83 EDMFKLATHPPKRVAVCFEVAQSRSDKKTCTMKEGHPGPFWDQHFVSNKSELF 142
DB 61 EDMFKLATHPPKRVAVCFEVAQSRSDKKTCTMKEGHPGPFWDQHFVSNKSELF 120
QY 143 TGISFSASVREQMSORFSPKEHPVLALPGAPAFQFVLEHRPLQKYMWSDEMVKTGEAQ 202
DB 121 TGISFSASVREQMSORFSPKEHPVLALPGAPAFQFVLEHRPLQKYMWSDEMVKTGEAQ 180
QY 203 IHAHLVRPVGIIHLRIGSDWKNACMLKDGTSAGSHFMASPCQVGYSRSTAAPLTMTCILP 262
DB 181 IHAHLVRPVGIIHLRIGSDWKNACMLKDGTSAGSHFMASPCQVGYSRSTAAPLTMTCILP 240
QY 263 DLKEIQRAVKLWVRSLDAQSYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILG 322
DB 241 DLKEIQRAVKLWVRSLDAQSYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILG 300
QY 323 QADHFIQNCVSSFTAFVKRERDLQGRSSFFGMDRPPKLRDEF 365
DB 301 QADHFIQNCVSSFTAFVKRERDLQGRSSFFGMDRPPKLRDEF 343

RESULT 7
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45921
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-45921

Query Match      42.3%; Score 837; DB 4; Length 399;
Best Local Similarity 44.9%; Pred. No. 1.9e-85;
Matches 168; Conservative 62; Mismatches 118; Indels 26; Gaps 9;

QY 7 DPAGYLLYCPGCRGNQADHFLGSLAFKALNRTLVAVPWIEYQHHKPPFTNLHVSYQK 66
DB 25 DPGYLYTCPCMGFRGNQADHFLGSLAFKALNRTLVAVPWIEYR--RGEIURSRQVFFMT 82
QY 67 YFKLEPLQAYHRVISLEDFMFKLATHPPKRVAVCFE-----VAQSRSDKKTCTMKE 121
DB 83 YFEVEPLKEYHRVITMAQDFMHLADDIWPESERSVFCYKERYSLQOEKNDDPKPCHAKD 142
QY 122 GNPFGPFPQDFHVSFNKSELTGTSFSASYREQ---WSQRFSPKPEHPVLALPGAPAFQPV 178
DB 143 GNPFGPFPMTDFHDFVRSEFVAPLHFDVHNSAANKWQTKYPAESYVPLAFTGAPASFV 202
QY 179 LEEHRPLQKYMWSDEMVKTGEAQIHAHLVR--PYVGIHLRIGSDWKNACMLKDGTSAGSH 237
DB 203 QLENCKLQRYLWORSYREASKDFIREQLPRGAFGLGIHLRNGIDWVRACEHVKD---SQH 259
QY 238 FMASPCQVGYSRSTAAPLTMTCILPDLKETQRAVKLWVRSL-----DAQSYVATDSSE 290
```

```
DB 260 LFASPCQLGYKNRGA-LYPELQCMPSKEAIIRLQKRTIKNVRQTPDNEIKSVFVSDSN 318
QY 291 SYVPELQ-QLFKRGKVKVSVLKPVAQVDLYILGQADHFIQNCVSSFTAFVKRERDLQGRP 349
DB 319 HMIQELNTALSRMIGISVHKLTEDDPYLDLAILGQSNHFIQNCISSYSAFVKRERDVHGFP 378
QY 350 SSFFGMDRPPKLRD 363
DB 379 SYFWGF---PKEK 389

RESULT 8
US-09-978-741-8
; Sequence 8, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-Nov-2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31-
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-8

Query Match      27.4%; Score 541.5; DB 3; Length 474;
Best Local Similarity 36.2%; Pred. No. 5.6e-52;
Matches 125; Conservative 60; Mismatches 107; Indels 53; Gaps 11;

QY 19 GRFGNQADHFLGSLAFKALNRTLVAVPWIEYQHHKPPFTNLHVSYQYFKLEPLQAYHR 78
DB 166 GRFGNQVDQLGLVLAFAKALDRTLVLPNFIEFKH--PETKM-IPFEFLFOVGTVAKYTR 221
QY 79 VISLEDPMFKLATHPPKRVAVCFEVAQSRSDKKTCTMKEGHPGPFWDQHFVSNK 138
DB 222 VVTMQEFTKIMTFHVGTPPROA-IYDKSAEPG-----CHSRKGNPFGFPWDQIDVSEVG 275
QY 139 SELFTGI--SFSASY---REQMSORFSPKPEHPVLALPGAPAFQVLEHRPLQKYMWS 193
DB 276 DEYFGDIPGGFDLQNGSKKWLKEKFSEYVPLAFSSAPAPPFGSKGKWSIQKYLWSS 335
QY 194 EMVKTGBAQIHAHLVRPYVGIHLRIGSDWKNACMLKDGTSAGSHFMASPCQVGYSRSTA 253
DB 336 RITEQAQKFIISANLAKFPFVAVHLRNDADWVRVCEHT-DTTTNRPLFASEQCLGEGHHLGT 394
QY 254 PLTMTMCLPDLKEIQRAVKLWVRSLDAQSYVATDSSEYVPELQQLFKGKVKVSLKPEV 313
```

```
Db 395 -ITKEICSPSKQOI-----LEQ-----IEAHROEPDD 420
Qy 314 AQVDLYLGOADHFIGNCVSFTAFVKRERDLOG---RPSSFFGM 355
Db 421 MYTSLAIMGRADLFVGNVCVSTFESHIVKRRDHAGQSPRPSAFFGI 465

RESULT 9
US-09-333-729A-12
; Sequence 12, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 12
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Caenorhabditis Elegans
US-09-333-729A-12

Query Match 27.4%; Score 541.5; DB 3; Length 474;
Best Local Similarity 36.2%; Pred. No. 5.6e-52;
Matches 125; Conservative 60; Mismatches 107; Indels 53; Gaps 11;

Qy 19 GREGNQADHFLGSLAFKLNRTLAVPPWTEYQHHKPPPTNLHVSKYKFKLEPLQAYHR 78
Db 166 GREGNQVDQFLGVLAFAKALDRTLVLPNFTFEXH---PETKM-IPFEFLFQVGTVAKYTR 221

Qy 79 VISLEDFMEKLAHTHPPEKRYVAYCEVAQORSPDKKTCPMKEGNEPGRPEWDQFHYSENK 138
Db 222 VVTMQEFTKIMTFHFVGTFRQA-IYDKSABPG-----CHSKEGNPPGPDWDQIDVSFVG 275

Qy 139 SELFTGI--SFSASY--RQWSEORFSPKEHPVLALPGAPAPQFVLEHRPLQKYMVWSD 193
Db 276 DEFYGDIPGFDLQMGSRKWKLEKFPSEYFVLAFSSAPAPFSGKVSQIYKLWSS 335

Qy 194 EMVKTBGAQIAHLVRPVIYVGIHRIKSDWKNVACAMLKDGTAGSHFMAPOCVGYSRSTAA 253
Db 336 RITEQAKKFISANLAKFPFVAHLRNDADWVRVCEHI-DTITNRPFLFASQCLGEGHHLGT 394

Qy 254 PLTWMTCLPDLKRIQRAVKLVWRSLDQSVYVATDSYVPELQQLFKGKVVVSLKPEV 313
Db 395 -LTKEICSPSKQOI-----LEQ-----IEAHROEPDD 420

Qy 314 AQVDLYLGOADHFIGNCVSFTAFVKRERDLOG---RPSSFFGM 355
Db 421 MYTSLAIMGRADLFVGNVCVSTFESHIVKRRDHAGQSPRPSAFFGI 465
```

1-61 of 2

```
RESULT 10
US-08-978-741-9
; Sequence 9, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NO. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-9

Query Match 17.8%; Score 353; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 4e-32;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLNRTLAVPPWIEYQHHKPPPTNL 60
Db 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLNRTLAVPPWIEYQHHKPPPTNL 60

Qy 61 H 61
Db 61 H 61

RESULT 11
US-09-333-729A-4
; Sequence 4, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 4
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-333-729A-4

Query Match 17.8%; Score 353; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 4e-32;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLNRTLAVPPWIEYQHHKPPPTNL 60
Db 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFKLNRTLAVPPWIEYQHHKPPPTNL 60

Qy 61 H 61
Db 61 H 61

RESULT 12
US-09-270-767-61459
```



```

; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
US-08-978-741-14

```

```

Query Match      5.1%; Score 101; DB 3; Length 28;
Best Local Similarity 90.0%; Pred. No. 0.00031;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 MPAGSWDPAGLYLPCMG 20
Db 9 MPAGSWDPAGLYLPCMG 28

```

```

Search completed: October 25, 2005, 15:42:47
Job time : 60.9589 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 232.341 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLPCMGK.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 16Dec04:.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	365	2 AAW80571	Aaw80571 Human hea
2	1979	100.0	388	8 ABM82223	Abm82223 Tumour-as
3	1979	100.0	397	2 AAW80573	Aaw80573 Human O-f
4	1867.5	94.4	417	4 AAM78377	Aam78377 Human pro
5	1855.5	93.8	417	4 AAM79361	Aam79361 Human pro
6	1844	93.2	343	2 AAW80577	Aaw80577 Partial h
7	831	42.0	402	4 ABB63790	Abb63790 Drosophil
8	809	40.9	165	4 AAE03110	Aae03110 Human gen
9	809	40.9	165	5 ABG63619	Abg63619 Human alb
10	809	40.9	165	8 ADL76884	Adl76884 Albumin f
11	353	17.8	61	2 AAW80574	Aaw80574 N-termina
12	295	14.9	61	2 AAW80572	Aaw80572 N-termina
13	164.5	8.3	328	5 ABG95669	Abg95669 Human nuc
14	159.5	8.1	490	4 ABB69800	Abb69800 Drosophil
15	159.5	8.1	490	8 ADS96504	Ads96504 Drosophil
16	142	7.2	429	5 ABG95651	Abg95651 Human nuc
17	120	6.1	150	3 AAY87325	Aay87325 Human sig
18	105.5	5.3	428	3 AAB40332	Aab40332 Human ORF
19	104	5.3	566	3 AAG55221	Aag55221 Arabidops
20	102.5	5.2	260	6 ABJ25654	Abj25654 Aspergill
21	102	5.2	566	3 AAG60442	Aag60442 Arabidops
22	101.5	5.1	783	1 AAP94265	Aap94265 Sequence
23	101.5	5.1	28	2 AAW80575	Aaw80575 N-termina
24	100.5	5.1	772	8 ABM84900	Abm84900 Human dia
25	100.5	5.1	772	8 ABM84902	Abm84902 Human dia

26	100.5	5.1	772	8	ABM84903	Abm84903 Human dia
27	100.5	5.1	772	8	ABM84899	Abm84899 Human dia
28	100	5.1	207	7	AD955590	Ad955590 Human NOV
29	100	5.1	207	7	AD955588	Ad955588 Human NOV
30	98.5	5.0	1216	3	ABA42097	Aba42097 Human ORF
31	94.5	4.8	349	6	ADA21127	Ada21127 Human sec
32	94.5	4.8	764	8	ABM84901	Abm84901 Human dia
33	94.5	4.8	4773	7	ADJ95094	Adj95094 Novel NOV
34	94	4.7	1449	7	ADC71275	Adc71275 Human col
35	93	4.7	692	3	AB49313	Ab49313 Mouse bro
36	92.5	4.7	1251	6	ABU35306	Abu35306 Protein e
37	92.5	4.7	1271	8	ADL06054	Adl06054 M. catarr
38	92.5	4.7	2074	3	AA543119	Aay43119 Amino aci
39	92	4.6	348	6	ABU25276	Abu25276 Protein e
40	92	4.6	650	6	ABU45346	Abu45346 Protein e
41	91.5	4.6	344	7	ADB97786	Adb97786 Human alp
42	91	4.6	1381	5	ABG93234	Abg93234 C. albica
43	91	4.6	1381	5	ABP73947	Abp73947 Candida a
44	90.5	4.6	353	7	AD955586	Ad955586 Human NOV
45	90.5	4.6	455	4	ABG25040	Abg25040 Novel hum

ALIGNMENTS

RESULT 1
AAW80571
ID AAW80571 standard; protein; 365 AA.
XX
AC AAW80571;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..61
FT /note= "N-terminal sequence claimed for in claim 4"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR N-PSDB; AAV65632.
XX
PT Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Fig 12A; 90pp; English.
XX
CC This represents a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EGF domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC Probes based on EGF domain polypeptide are used to detect gene

CC amplification and expression. The expression can also be determined at
CC the protein level using antibodies specific for O-fucosyltransferase
XX
SQ Sequence 365 AA;

Query Match 100.0%; Score 1979; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.5e-205; Indels 0; Gaps 0;
Matches 365; Conservative 0; Mismatches 0;

QY 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPFTNL 60
DB 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPFTNL 60

QY 61 HVSQYQYFKLEPQYHVRVLSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQSPDKTCPMK 120
DB 61 HVSQYQYFKLEPQYHVRVLSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQSPDKTCPMK 120

QY 121 EGNPFGPFDQPHVSNFNSKSELTGTISFSASYREQWSQSFSPKEHPVLALPCAPQFPVLE 180
DB 121 EGNPFGPFDQPHVSNFNSKSELTGTISFSASYREQWSQSFSPKEHPVLALPCAPQFPVLE 180

QY 181 EHRPQKYMVMSDEWVKTEGAQIHAHLVRPYYGIIHLRIGSDWKACAMLDKGTAGSHFMA 240
DB 181 EHRPQKYMVMSDEWVKTEGAQIHAHLVRPYYGIIHLRIGSDWKACAMLDKGTAGSHFMA 240

QY 241 SPQCVGYSRSTAAPLTMTMCLPDKEIQRAVKLWVRSLDAQSVVATDSSEYVPELQQLF 300
DB 241 SPQCVGYSRSTAAPLTMTMCLPDKEIQRAVKLWVRSLDAQSVVATDSSEYVPELQQLF 300

QY 301 KGKVKVYSLKPEVAQVDLYILQADHFIGNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
DB 301 KGKVKVYSLKPEVAQVDLYILQADHFIGNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360

QY 361 LRDEF 365
DB 361 LRDEF 365

RESULT 2
ID AEW82223
AC AEW82223;
XX
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
FN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
DR N-PSDB; ACN40746.
XX

PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 5726; 7273pp; English.

PS The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT polypeptide of the invention
XX
SQ Sequence 388 AA;

Query Match 100.0%; Score 1979; DB 8; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.7e-205; Indels 0; Gaps 0;
Matches 365; Conservative 0; Mismatches 0;

QY 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPFTNL 60
DB 24 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPFTNL 83

QY 61 HVSQYQYFKLEPQYHVRVLSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQSPDKTCPMK 120
DB 84 HVSQYQYFKLEPQYHVRVLSLEDFMEKLAPTHWPPEKRVAYCFEVAQAQSPDKTCPMK 143

QY 121 EGNPFGPFDQPHVSNFNSKSELTGTISFSASYREQWSQSFSPKEHPVLALPCAPQFPVLE 180
DB 144 EGNPFGPFDQPHVSNFNSKSELTGTISFSASYREQWSQSFSPKEHPVLALPCAPQFPVLE 203

QY 181 EHRPQKYMVMSDEWVKTEGAQIHAHLVRPYYGIIHLRIGSDWKACAMLDKGTAGSHFMA 240
DB 204 EHRPQKYMVMSDEWVKTEGAQIHAHLVRPYYGIIHLRIGSDWKACAMLDKGTAGSHFMA 263

QY 241 SPQCVGYSRSTAAPLTMTMCLPDKEIQRAVKLWVRSLDAQSVVATDSSEYVPELQQLF 300
DB 264 SPQCVGYSRSTAAPLTMTMCLPDKEIQRAVKLWVRSLDAQSVVATDSSEYVPELQQLF 323

QY 301 KGKVKVYSLKPEVAQVDLYILQADHFIGNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 360
DB 324 KGKVKVYSLKPEVAQVDLYILQADHFIGNCVSSFTAFVKRERDLQGRPSFFGMDRPPK 383

QY 361 LRDEF 365
DB 384 LRDEF 388

RESULT 3
AAW80573
ID AAW80573 standard; protein; 397 AA.
XX
XX AAW80573;
XX
XX 16-DEC-1998 (first entry)
DT
XX Human O-fucosyltransferase sequence expressed by a plasmid insertion.
DE
XX

KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
 KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
 OS Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Peptide 27..32
 FT /note= "polyhistidine tag"

XX WO9833924-A1.

XX 06-AUG-1998.

XX 17-DEC-1997; 97WO-US023401.

XX 31-JAN-1997; 97US-00792498.

XX 26-NOV-1997; 97US-00978741.

XX (GETH) GENENTECH INC.

XX Wang Y, Spellman MW;
 XX WPI; 1998-437477/37.
 XX N-PSDB; AAV65633.

XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.
 XX Example; Fig 13B; 90pp; English.

XX This represents the amino acid sequence of the human heart O-
 CC fucosyltransferase expressed by a plasmid insertion. The human O-
 CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
 CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
 CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
 CC domains, are used in diagnosis and treatment of conditions associated
 CC with overexpression of O-fucosyltransferase, to promote survival of
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
 CC used to detect gene amplification and expression. The expression can also
 CC be determined at the protein level using antibodies specific for O-
 CC fucosyltransferase

XX Sequence 397 AA;

Query Match 100.0%; Score 1979; DB 2; Length 397;
 Best Local Similarity 100.0%; Pred. No. 2.8e-205;
 Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGYLLYPCWGRFGNQADHFLGSLAPAKLLNRTLAVPPWIEYQHHKPPFTNL 60
 Db 33 MPAGSWDPAGYLLYPCWGRFGNQADHFLGSLAPAKLLNRTLAVPPWIEYQHHKPPFTNL 92
 Qy 61 HVSQYKFLPELQAHVRLSLEDFMEKLAFTWPPKRVAYCFEVAQAQSPDKKTCPEK 120
 Db 93 HVSQYKFLPELQAHVRLSLEDFMEKLAFTWPPKRVAYCFEVAQAQSPDKKTCPEK 152
 Qy 121 EGNPFGFQDFHVSFNKSLFTGISFSASREQWSQFSPKHPVLALPGAPAFVLE 180
 Db 153 EGNPFGFQDFHVSFNKSLFTGISFSASREQWSQFSPKHPVLALPGAPAFVLE 212
 Qy 181 EHRPLQKYMVMSDEWMTGTEAQAHAHLVVRPVVGHILRIGSDWKNACAMLDKGTAGSHFMA 240
 Db 213 EHRPLQKYMVMSDEWMTGTEAQAHAHLVVRPVVGHILRIGSDWKNACAMLDKGTAGSHFMA 272
 Qy 241 SPQCQVGSRSRTAAELTWTMCLPDLEKEIQRVAVKLVRSILDAQSVVYVATDSYVPELQOLF 300
 Db 273 SPQCQVGSRSRTAAELTWTMCLPDLEKEIQRVAVKLVRSILDAQSVVYVATDSYVPELQOLF 332
 Qy 301 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKKRERDLOQRPSFFGMDRPPK 360
 Db 333 KGKVKVSLKPEVAQVDLYILGQADHFIQNCVSSFTAFVKKRERDLOQRPSFFGMDRPPK 392

Qy 361 LRDEF 365
 Db 393 LRDEF 397

RESULT 4
 AAM78377

ID AAM78377 standard; protein; 417 AA.

XX AC AAM78377;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1039.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX N-PSDB; AAK51510.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.

XX Claim 20; Page 3276-3277; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibit activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM00020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX SQ Sequence 417 AA;

Query Match 94.4%; Score 1867.5; DB 4; Length 417;
 Best Local Similarity 96.2%; Pred. No. 3.8e-193;
 Matches 350; Conservative 3; Mismatches 6; Indels 5; Gaps 1;

Qy 7 DPAGYLLYPCW-----GREGNQADHFLGSLAPAKLLNRTLAVPPWIEYQHHKPPFTNLH 61

Db. 54 DMATWTHAPSEVEGGQGRFGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHPPTNLH 113
QY 62 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAQQRSPDKTCPMKE 121
Db 114 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAQQRSPDKTCPMKE 173
QY 122 GNPFGPFMDQHFVSNKSELFTGISFSASYREQWSQFSPKEHPVLAALPGAPAQFPVLEE 181
Db 174 GNPFGPFMDQHFVSNKSELFTGISFSASYREQWSQFSPKEHPVLAALPGAPAQFPVLEE 233
QY 182 HRPLQKYVMSDEWVKTCGEAQIHAHLVRPVGVIHLRIGSDWKACAMLDKGTAGSHFMA 241
Db 234 HRPLQKYVMSDEWVKTCGEAQIHAHLVRPVGVIHLRIGSDWKACAMLDKGTAGSHFMA 293
QY 242 PQCVGYSRSTAAPLTMTMCLPDLKEIORAVKLVWRSLDQSVYVATDSSEYVPELQOLF 301
Db 294 PQCVGYSRSTAAPLTMTMCLPDLKEIORAVKLVWRSLDQSVYVATDSSEYVPELQOLF 353
QY 302 GKVKVYSLKPEVAQVDLYILQADHFTGNCVSSFTAFVKRERDLQGRPSFFGMDRPPKL 361
Db 354 GKVKVYSLKPEVAQVDLYILQADHFTGNCVSSFTAFVKRERDLQGRPSFFGMDRPPKL 413
QY 362 RDEF 365
Db 414 RDEF 417

RESULT 5
AAW79361
ID AAW79361 standard; protein; 417 AA.
XX
AC AAW79361;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3007.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
FN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX

(HYSE-) HYSEQ INC.
XX
PI Tang Y, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52494.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
XX Claim 20; Page 229-230; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAW78223-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The vaccines or
CC polynucleotides and polypeptides are useful in gene therapy, activities,
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin acter, and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 417 AA;
Query Match 93.8%; Score 1855.5; DB 4; Length 417;
Best Local Similarity 95.6%; Fred. No. 7.6e-192;
Matches 348; Conservative 4; Mismatches 7; Indels 5; Gaps 1;
QY 7 DPAGYLLYCPM-----GRFGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHPPTNLH 61
Db 54 DMATWTHAPSEVEGGQGRFGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHPPTNLH 113
QY 62 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAQQRSPDKTCPMKE 121
Db 114 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAQQRSPDKTCPMKE 173
QY 122 GNPFGPFMDQHFVSNKSELFTGISFSASYREQWSQFSPKEHPVLAALPGAPAQFPVLEE 181
Db 174 GNPFGPFMDQHFVSNKSELFTGISFSASYREQWSQFSPKEHPVLAALPGAPAQFPVLEE 233
QY 182 HRPLQKYVMSDEWVKTCGEAQIHAHLVRPVGVIHLRIGSDWKACAMLDKGTAGSHFMA 241
Db 234 HRPLQKYVMSDEWVKTCGEAQIHAHLVRPVGVIHLRIGSDWKACAMLDKGTAGSHFMA 293
QY 242 PQCVGYSRSTAAPLTMTMCLPDLKEIORAVKLVWRSLDQSVYVATDSSEYVPELQOLF 301
Db 294 PQCVGYSRSTAAPLTMTMCLPDLKEIORAVKLVWRSLDQSVYVATDSSEYVPELQOLF 353
QY 302 GKVKVYSLKPEVAQVDLYILQADHFTGNCVSSFTAFVKRERDLQGRPSFFGMDRPPKL 361
Db 354 GKVKVYSLKPEVAQVDLYILQADHFTGNCVSSFTAFVKRERDLQGRPSFFGMDRPPKL 413
QY 362 RDEF 365
Db 414 RDEF 417

RESULT 6
AAW80577
ID AAW80577 standard; protein; 343 AA.
XX
AC AAW80577;
XX
DT 16-DEC-1998 (first entry)
XX
DE Partial human sequence of unknown function from a myeloblast cell line.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX
FN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
XX 17-DEC-1997; 97WO-US023401.
XX
XX 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00578741.
PR

XX PA (GETH) GENENTECH INC.
XX PI Wang Y, Spellman MW;
XX DR WPI; 1998-437477/37.
XX DR N-PSDB; AAV65634.
XX PT Human O-fucosyltransferase able to glycosylate epidermal growth factor
XX PT overexpression of the enzyme.
XX PS Example 1; Page 62-68; 90pp; English.
XX CC This represents a published partial human sequence of unknown function
XX CC from a myeloblast cell line. The invention provides a human heart O-
XX CC fucosyltransferase enzyme that can glycosylate an epidermal growth factor
XX CC (EGF) domain of a polypeptide with an activated O-fucose residue.
XX CC Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
XX CC for the EGF domains, are used in diagnosis and treatment of conditions
XX CC associated with overexpression of O-fucosyltransferase, to promote
XX CC survival of sensory (retinal) neurons. Probes based on EGF domain
XX CC polypeptide are used to detect gene amplification and expression. The
XX CC expression can also be determined at the protein level using antibodies
XX CC specific for O-fucosyltransferase
XX SQ Sequence 343 AA;
Query Match 93.2%; Score 1844; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 1e-190;
Matches 343; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 NQADHFLGSLAFKALNRTLAVPPWIEYQHKKPPFTNLHVSQYKFLPLQAYHVRVISL 82
Db 1 NQADHFLGSLAFKALNRTLAVPPWIEYQHKKPPFTNLHVSQYKFLPLQAYHVRVISL 60
QY 83 EDPMFKLAPTHWPEKRVAYCFEVAQAORSDDKTCPMKEGNPFPWDQPHVFNKSELF 142
Db 61 EDPMFKLAPTHWPEKRVAYCFEVAQAORSDDKTCPMKEGNPFPWDQPHVFNKSELF 120
QY 143 TGISFSASYEQNSQSPKEHPVLALPGAPAPFPVLEHRPLQKYMVSDENVKTEGAQ 202
Db 121 TGISFSASYEQNSQSPKEHPVLALPGAPAPFPVLEHRPLQKYMVSDENVKTEGAQ 180
QY 203 IHAHLVRPYYGIIHLRIGSDWKNAACAMLDKGTAGSHFMASSPCQVGYRSRTAAPLTMTWCLP 262
Db 181 IHAHLVRPYYGIIHLRIGSDWKNAACAMLDKGTAGSHFMASSPCQVGYRSRTAAPLTMTWCLP 240
QY 263 DLKEIQRAVKLWRSLLDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYLG 322
Db 241 DLKEIQRAVKLWRSLLDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYLG 300
QY 323 QADHFIGNCVSSFTAFVKRERDLQGRSSPFGMDRPPKLRDEF 365
Db 301 QADHFIGNCVSSFTAFVKRERDLQGRSSPFGMDRPPKLRDEF 343
RESULT 7
ABB63790
ID ABB63790 standard; protein; 402 AA.
XX AC ABB63790;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 18162.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX XX WO200171042-A2.
XX FN

XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX XX 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX XX (PEKE) PE CORP NY.
XX PA Venter JC, Adams M, Li PWD, Myers EW;
XX PI N-PSDB; ABL07893.
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL07893.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 402 AA;
Query Match 42.0%; Score 831; DB 4; Length 402;
Best Local Similarity 44.7%; Pred. No. 1.1e-80;
Matches 167; Conservative 62; Mismatches 119; Indels 26; Gaps 9;
QY 7 DPAGYLLYPCMGRCFGNQADHFLGSLAFKALNRTLAVPPWIEYQHKKPPFTNLHVSQYK 66
Db 28 DPNGYLLYPCMGRCFGNQADHFLGSLAFKALNRTLILPPWVEYR--RGELRSRQVFPNT 85
QY 67 YFLEPLQAYHVRVISLEDFMEKLAPTHWPEKRVAYCFE-----VAAQSPDKKTCPMKE 121
Db 86 YFEVEPLKEVHRVITWADFMWHLADDIWPESERVFCYKERYSLQOEKNPDKPNCHAKD 145
QY 122 GNPFGPFWDFHVSFNKSELFTGISFSASYREQ---WSQRFSPKEHPVLALPGAPAPFPV 178
Db 146 GNPFGPFWDTFHIDFVRSEFYAPLHFDVHHSNEAAKMQTKYPAESYVPLAFTGAPASFPV 205
QY 179 LEEHRPLQKYMVSDENVKTEGAQIHAHLVR-PVVGIIHLRIGSDWKNAACAMLDKGTAGSH 237
Db 206 QLENCCKLQRYLQWSQRYREASKDFIREQLPRGAPLGIHLRNGIDWVRACHEVKD---SQH 262
QY 238 FMASPCQVGYRSRTAAPLTMTWCLPLDKETIQRAVKLWRSLL-----DAQSVYVATDS 290
Db 263 LFASPCQLGYKNERGA-LYPELCWPSKEAIIQLKRTIKNVROTQPDNEIKSVFVSDSN 321
QY 291 SYYPELQ-QLFKGKVKVSLKPEVAQVDLYLGQADHFIGNCVSSFTAFVKRERDLQGRP 349
Db 322 HMIGELNTALSRMGISVHKLPEDDPVLDLAILGOSNHFIGNCISYSAFKEKRDVHGFP 381
QY 350 SSPFGMDRPPKLRD 363
Db 382 SYFWGF---PKEKD 392
RESULT 8
AAE03110
ID AAE03110 standard; protein; 165 AA.
XX AC AAE03110;
XX XX

DT 10-AUG-2001 (first entry)
DE Human gene 18 encoded secreted protein HMWEY26, SEQ ID NO:144.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
KW chemotaxis; food additive; chromosome 3; binding partner identification.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..1 /label= Signal_peptide
FT Protein 2..165 /note= "Mature secreted protein"
FT Misc-difference 2 /label= Unknown
FT /note= "Encoded by ANA"
FT Misc-difference 4 /label= Unknown
FT /note= "Encoded by NCA"
FT Misc-difference 5 /label= Unknown
FT /note= "Encoded by TGA"
FT Misc-difference 148 /label= Unknown
FT /note= "Encoded by AGS"
XX
PN WO200132676-A1.
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US029365.
XX
PR 29-OCT-1999; 99US-0162237P.
PR 21-JUL-2000; 2000US-0219666P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Shi Y, Olsen HS, Soppet DR;
PI N-PSDB; AAD07629.
DR WPI; 2001-328773/34.
DR N-PSDB; AAD07629.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT diabetes mellitus and multiple sclerosis.
XX
PS Claim 11; Page 486-487; 546pp; English.
XX
CC AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g. by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 25 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin

CC disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention
XX
SQ Sequence 165 AA;

Query Match 40.9%; Score 809; DB 4; Length 165;
Best Local Similarity 96.9%; Pred. No. 6.6e-79;
Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 206 HLVRPVYVGIHLRIGSDWKNACAMLKDGTFAGSHFMASPCQVGYSRSTAAPLMTWCLPDLK 265
Db 6 HFARTYPGIHLRIGSDWKNACAMLKDGTFAGSHFMASPCQVGYSRSTAAPLMTWCLPDLK 65

QY 266 EIQRVKLVWRSIDAQSVVATDSESVPELOQLFKGVKVVSLKPEVAQVDLYILGQAD 325
Db 66 EIQRVKLVWRSIDAQSVVATDSESVPELOQLFKGVKVVSLKPEVAQVDLYILGQAD 125

QY 326 HFTGNCVSSFTAFVXRERDLQGRPSFFGMDRPPKLRDEF 365
Db 126 HFTGNCVSSFTAFVXRERDLQGRPSFFGMDRPPKLRDEF 165

RESULT 9
ABG63619
ID ABG63619 standard; protein; 165 AA.
XX
AC ABG63619;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human albumin fusion protein #294.
XX
KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW human serum albumin; HSA; cancer; reproductive disorder;
KW digestive disorder; immune disorder; endocrine disorder;
KW haematopoietic disorder; neural disorder; connective disorder;
KW cytostatic; antiinfertility; antiinflammatory; antiulcer;
KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW osteopathic; antiarthritic.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200177137-A1.
XX
PD 18-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US011988.
XX
PR 12-APR-2000; 2000US-0229358P.
PR 25-APR-2000; 2000US-0199384P.
PR 21-DEC-2000; 2000US-0256931P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Haseltine WA;
XX WPI; 2002-010886/01.
XX
PT New fusion protein for treating disease e.g. diabetes comprises an

PT albumin fused to a therapeutic protein.
XX Claim 1; Page 755-756; 2102pp; English.
XX

CC The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or disorder
CC that may be modulated by therapeutic protein X. The albumin extends the
CC shelf-life of protein X, and may increase its biological in vitro/in vivo
CC activity. The protein is useful for treating and diagnosing disorders
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
CC disease, ulcerative colitis), immune disorders (e.g. acquired
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
CC schizophrenia), and connective tissue disorders (e.g. osteoporosis, arthritis).
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
XX Sequence 165 AA;

Query Match 40.9%; Score 809; DB 5; Length 165;
Best Local Similarity 96.9%; Pred. No. 6.6e-79;
Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 206 HLVRPVYGIHLRIGSDWKNACAMLDGTAGSHFMAWPQCVCYGRSTAAPLTMTCLPDLK 265
DB 6 HFARTYPIHILRIGSDWKNACAMLDGTAGSHFMAWPQCVCYGRSTAAPLTMTCLPDLK 65

QY 266 EIQRVAVKLVRSLSDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILQAD 325
DB 66 EIQRVAVKLVRSLSDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILQAD 125

QY 326 HFIGNCVSSFTAFVKRERDIQGRPSFFGMDRPPKLRDEF 365
DB 126 HFIGNCVSSFTAFVKRERDIQGRPSFFGMDRPPKLRDEF 165

RESULT 10
ADL76884 standard; protein; 165 AA.

XX ADL76884;
XX
XX 20-MAY-2004 (first entry)
XX
XX Albumin fusion protein related therapeutic protein X, SEQ ID No 366.
XX
XX albumin fusion protein; cytostatic; antianaemic; antiarthritic;
XX antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
XX antipsoriatic; antibacterial; osteopathic; dermatological; antigout;
XX immunomodulator; antiarrhythmic; cardiac; nootropic; antilipemic;
XX nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
XX antidiabetic; anabolic; hypertensive; vulnerable; gene therapy; cancer;
XX reproductive system disorder; therapeutic protein.

Unidentified.

US2004010134-A1.

15-JAN-2004.

12-APR-2001; 2001US-00833245.

12-APR-2000; 2000US-0229358P.

25-APR-2000; 2000US-0199384P.

21-DEC-2000; 2000US-0256931P.

(ROSE/) ROSEN C A.

(HASE/) HASELTINE W A.

Rosen CA, Haseltine WA;

WPI; 2004-090519/09.

XX New albumin fusion proteins, useful for diagnosing, treating, preventing
PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,
PT asthma, inflammatory bowel disease or Alzheimer's disease.
XX Disclosure; SEQ ID NO 366; 279pp; English.

XX The invention relates to a novel albumin fusion protein. The invention
further relates to: a composition comprising the albumin fusion protein
and a pharmaceutical carrier; a kit comprising the composition of the
albumin fusion protein formula; a method of treating a disease or
disorder in a patient comprising the step of administering the albumin
fusion protein; a method of treating a patient with a disease or disorder
that is modulated by Therapeutic protein: X, or its fragment or variant;
a method of extending the shelf life of Therapeutic protein: X, or its
fragment or variant; a nucleic acid molecule comprising a polynucleotide
sequence encoding the albumin fusion protein; a vector comprising the
nucleic acid molecule of the albumin fusion protein; and a host cell
comprising the nucleic acid molecule of the albumin fusion protein. The
albumin fusion protein and its compositions have the following
activities: cytostatic, antianaemic, antiarthritic, antiasthmatic, anti-
HIV, immunosuppressive, antiinflammatory, antipsoriatic, antibacterial,
osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic,
cardiac, nootropic, antilipemic, nephrotropic, uropathic,
neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
hypertensive, and vulnerable. The albumin fusion protein nucleic acid may
be used in gene therapy to treat disorders. The albumin fusion protein is
useful for diagnosing, treating, preventing or ameliorating diseases or
disorders comprising indication: Y. The diseases or disorders include:
cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer), acute
immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, AIDS,
lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
disease), reproductive system disorders (e.g. prostatitis, inguinal
hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
Leydig tumours), musculoskeletal diseases (e.g. giant cell tumours, Paget's
disease, systemic lupus erythematosus, gout, muscular dystrophy or
cachexia), cardiovascular disease (e.g. rhabdomyoma, heart disease,
arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or
hyponatraemia), mixed foetal diseases (e.g. foetal alcohol syndrome,
Down's syndrome, Patau syndrome, Turner's syndrome, Apert syndrome or Tay
-Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
tract infections or renal disorders), neural or sensory disease (e.g.
Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
cerebellar ataxia, attention deficit disorder, autism or obsessive
compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
disease or glomerulonephritis), digestive diseases (e.g. portal
hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
or connective tissue or epithelial diseases (e.g. Crohn's disease,
scleroderma, wound healing or epidermolysis bullosa). This sequence
represents a therapeutic protein X relating to the albumin fusion protein
of the invention. The sequence listing data for this specification was
downloaded from the USPTO website.

XX Sequence 165 AA;

Query Match 40.9%; Score 809; DB 8; Length 165;

Best Local Similarity 96.9%; Pred. No. 6.6e-79;

Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 206 HLVRPVYGIHLRIGSDWKNACAMLDGTAGSHFMAWPQCVCYGRSTAAPLTMTCLPDLK 265
DB 6 HFARTYPIHILRIGSDWKNACAMLDGTAGSHFMAWPQCVCYGRSTAAPLTMTCLPDLK 65

QY 266 EIQRVAVKLVRSLSDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILQAD 325
DB 66 EIQRVAVKLVRSLSDAQSVYVATDSSEYVPELQQLFKGKVKVSLKPEVAQVDLYILQAD 125

QY 326 HFIGNCVSSFTAFVKRERDIQGRPSFFGMDRPPKLRDEF 365
DB 126 HFIGNCVSSFTAFVKRERDIQGRPSFFGMDRPPKLRDEF 165

RESULT 11
AAW80574
ID AAW80574 standard; peptide; 61 AA.
XX
AC AAW80574;
DT 16-DEC-1998 (first entry)
XX N-terminal amino acid sequence of human heart O-fucosyltransferase.
DE
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
KW
XX Homo sapiens.
OS
XX W09833924-A1.
FN
XX 06-AUG-1998.
PD
XX 17-DEC-1997; 97WO-US023401.
PF
XX 31-JAN-1997; 97US-00792498.
PR
XX 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
PA
XX Wang Y, Spellman MW;
PI
XX WPI; 1998-437477/37.
DR
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
XX Claim 4; Page 69; 90pp; English.
PS
XX This represents a the N-terminal sequence of the human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
XX Sequence 61 AA;
SQ

Query Match 17.8%; Score 353; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 4.4e-30; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCPCMGREGNOADHFLGSLAFAPAKLNLRTLVAPPWIEYOHKHPFTNL 60
DB 1 MPAGSWDPAGYLLYPCPCMGREGNOADHFLGSLAFAPAKLNLRTLVAPPWIEYOHKHPFTNL 60
QY 61 H 61
DB 61 H 61

RESULT 12
AAW80572
ID AAW80572 standard; peptide; 61 AA.
XX
AC AAW80572;
DT 16-DEC-1998 (first entry)
XX N-terminal amino acid sequence of CHO O-fucosyltransferase.
DE
XX

KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 15 /note= "unknown"
PT
FT Misc-difference 17 /note= "unknown"
FT
FT Misc-difference 38 /note= "unknown"
FT
XX W09833924-A1.
FN
XX 06-AUG-1998.
PD
XX 17-DEC-1997; 97WO-US023401.
PF
XX 31-JAN-1997; 97US-00792498.
PR
XX 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
PA
XX Wang Y, Spellman MW;
PI
XX WPI; 1998-437477/37.
DR
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
XX Claim 5; Page 69; 90pp; English.
PS
XX This represents the N-terminal sequence of CHO O-fucosyltransferase. The
CC enzyme can glycosylate an epidermal growth factor (EGF) domain of a
CC polypeptide with an activated O-fucose residue. Inhibitors of O-
CC fucosyltransferase, e.g. mutants with increased affinity for the EGF
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
XX Sequence 61 AA;
SQ

Query Match 14.9%; Score 295; DB 2; Length 61;
Best Local Similarity 91.5%; Pred. No. 8.6e-24; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3 AGSWDPAGYLLYPCPCMGREGNOADHFLGSLAFAPAKLNLRTLVAPPWIEYOHKHPFTNLH 61
DB 3 AGSWDLAGYLLYAPXMGREGNOADHFLGSLAFAPAKLXVRLVAPPWIEYOHKHPFTNLH 61
RESULT 13
ABG95669
ID ABG95669 standard; protein; 328 AA.
XX
XX ABG95669;
AC
XX
DT 06-DEC-2002 (first entry)
XX
XX Human nucleic acid associated protein, NAAP26, Incyte 7503717CD1.
DE
XX Human; nucleic acid-associated protein; NAAP; gene therapy; microarray;
KW proliferative disorder; developmental disorder; cardiovascular disorder;
KW neurological disorder; autoimmune disorder; inflammatory disorder;
KW atherosclerosis; bursitis; cirrhosis; hepatitis; psoriasis; cancer;
KW lymphoma; melanoma; brain cancer; breast cancer; Cushing's syndrome;
KW Alzheimer's disease; AIDS; Creutzfeldt-Jakob disease;
KW acquired immunodeficiency syndrome; Goodpasture's syndrome;
KW

CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 490 AA;
 Query Match 8.1%; Score 159.5; DB 4; Length 490;
 Best Local Similarity 22.4%; Pred. No. 1.1e-07;
 Matches 92; Conservative 55; Mismatches 159; Indels 105; Gaps 19;
 QY 9 AGVLLYCPQWGR-FGNOADHFLGSLAFKLLNR-----TLAVPPWIEYQH-HKPPFTN 59
 Db 71 AVILYDVNISEGFNLRDVIYRMVFRRLQRRFRHRLVLPFPRLYHWHSSQGLQ 130
 QY 60 LHVSYQKYFKLEPLOAHYHVISLEDFM--EKLAPTHWPKKRVAYCFEV-----YREQW 155
 Db 131 SGLPWFHFFDLASLRRYAPVLDYEEFLAEQRLFGNPGAPLVHVGHAFLQHYEVMLEQGI 190
 QY 107 ---AAQRSPDKKTCPMKEGN--PFGPFWDQPHVSFNKSELFTGISFSAS-----YREQW 155
 Db 191 FRDKFERVTDK---PCSEGLSGGFLQQAELRVGR--FHCVRFOGSAGLLEKLLREAI 244
 QY 156 SQRFSPKEH-----PVLALPGAPAPQFPVLEHRLPQKYMWSDVMVKTGEAQIHAHLVRPY 211
 Db 245 DEDTAGPEDVDDMRTYALLSAET---VLHDDH-----WGDE-----HFWQARRS 284
 QY 212 VGTHLR---IGSDWKNACAMLKDGTA-----GSHFMASPOCVGYSR 249
 Db 285 MRFARLEQVAADFRQALDITDASAGVQRPAMWELERPKRNAKG DYLCALHRRGDVFR 344
 QY 250 STAAPLTMTMCLPDLKEIQRVAVKLVRSIDAQSVYVATDSY-VPELOQLFKGKVKVVS 308
 Db 345 SRDA-----TTPTLKAQAQVQKLLRGFNMTTVFLATDTPYELMELKELFY-RFRLVH 397
 QY 309 LKPE-----VAQVDLYILGQADHFGNVCVSSFTAFVKRERDLOG 347
 Db 398 FAPESNVORRELKDGGA VVDQLVCAYARYFVGTYESTFTYRIYERBILG 448

RESULT 15
 ADS96504
 ID ADS96504 standard; protein; 490 AA.

AC ADS96504;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Drosophila melanogaster protein, SEQ ID 125.
 XX
 KW Insecticide; Antiparasitic; Antihelminthic.
 XX
 OS Drosophila melanogaster.
 XX
 EN WO2004039999-A2.
 XX
 PD 13-MAY-2004.

XX 08-AUG-2003; 2003WO-US024982.
 XX
 XX 30-OCT-2002; 2002US-0422377P.
 XX
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Stam L, Kamdar KP, Spana E, Bachmann J;
 XX WPI; 2004-376203/35.
 DR N-P5DB; ADS96503.
 XX

PT Identifying a compound that inhibits the activity of a protein for
 PT Drosophila viability for use e.g., as insecticidal agent by expressing in
 PT a recombinant host a DNA molecule to produce a protein essential for
 PT Drosophila viability.

XX Claim 1; SEQ ID NO 125; 57pp; English.

XX The present invention relates to a method for identifying a compound that
 CC inhibits the activity of a protein essential for Drosophila viability.
 CC The method comprises: (a) expressing in a recombinant host a DNA sequence
 CC encoding a protein essential for Drosophila viability; (b) testing
 CC compounds suspected of having the ability to inhibit the activity of the
 CC protein expressed in (a); and identifying a compound tested in (b) that
 CC inhibits the activity of the protein. The method is useful in identifying
 CC a compound that inhibits the activity of a protein essential for
 CC Drosophila viability for use as insecticidal, ectoparasiticidal,
 CC antiparasitic, antihelminthic or acaricidal agent. The present sequence
 CC is one such protein essential for Drosophila viability. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 490 AA;

Query Match 8.1%; Score 159.5; DB 8; Length 490;
 Best Local Similarity 22.4%; Pred. No. 1.1e-07;
 Matches 92; Conservative 55; Mismatches 159; Indels 105; Gaps 19;
 QY 9 AGVLLYCPQWGR-FGNOADHFLGSLAFKLLNR-----TLAVPPWIEYQH-HKPPFTN 59
 Db 71 AVILYDVNISEGFNLRDVIYRMVFRRLQRRFRHRLVLPFPRLYHWHSSQGLQ 130
 QY 60 LHVSYQKYFKLEPLOAHYHVISLEDFM--EKLAPTHWPKKRVAYCFEV-----YREQW 155
 Db 131 SGLPWFHFFDLASLRRYAPVLDYEEFLAEQRLFGNPGAPLVHVGHAFLQHYEVMLEQGI 190
 QY 107 ---AAQRSPDKKTCPMKEGN--PFGPFWDQPHVSFNKSELFTGISFSAS-----YREQW 155
 Db 191 FRDKFERVTDK---PCSEGLSGGFLQQAELRVGR--FHCVRFOGSAGLLEKLLREAI 244
 QY 156 SQRFSPKEH-----PVLALPGAPAPQFPVLEHRLPQKYMWSDVMVKTGEAQIHAHLVRPY 211
 Db 245 DEDTAGPEDVDDMRTYALLSAET---VLHDDH-----WGDE-----HFWQARRS 284
 QY 212 VGTHLR---IGSDWKNACAMLKDGTA-----GSHFMASPOCVGYSR 249
 Db 285 MRFARLEQVAADFRQALDITDASAGVQRPAMWELERPKRNAKG DYLCALHRRGDVFR 344
 QY 250 STAAPLTMTMCLPDLKEIQRVAVKLVRSIDAQSVYVATDSY-VPELOQLFKGKVKVVS 308
 Db 345 SRDA-----TTPTLKAQAQVQKLLRGFNMTTVFLATDTPYELMELKELFY-RFRLVH 397
 QY 309 LKPE-----VAQVDLYILGQADHFGNVCVSSFTAFVKRERDLOG 347
 Db 398 FAPESNVORRELKDGGA VVDQLVCAYARYFVGTYESTFTYRIYERBILG 448

Search completed: October 25, 2005, 15:30:37
 Job time : 236.341 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 08:18:39 ; Search time 320.031 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGLYLPCFMR.....QGRPSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09774954/runat 25102005_105433_6415/app_query.fasta_1.917
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN_1_116 @runat 25102005_105433_6415 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	1100	3	US-08-978-741-16
2	1979	100.0	1100	3	US-09-333-729A-16
3	1979	100.0	1300	3	US-08-978-741-4
4	1979	100.0	1300	3	US-09-333-729A-6
5	1979	100.0	1514	3	US-08-978-741-1
6	1979	100.0	1514	3	US-09-333-729A-2
7	1979	100.0	11284	3	US-08-978-741-5
8	1844	93.2	5009	3	US-08-978-741-7
9	1844	93.2	5009	3	US-09-333-729A-8
10	837	42.3	1320	4	US-09-270-767-14353
11	343	17.3	722	4	US-09-270-767-30507
12	186	9.4	998	4	US-09-270-767-28905

C 13	186	9.4	2034	4	US-09-270-767-13022	Sequence 13022, A
C 14	120	6.1	19227	4	US-09-949-016-12127	Sequence 12127, A
C 15	120	6.1	19228	4	US-09-949-016-16285	Sequence 16285, A
C 16	100.5	5.1	10259	4	US-09-949-016-13785	Sequence 13785, A
C 17	97.5	4.9	24741	4	US-09-949-016-15547	Sequence 15547, A
C 18	96.5	4.9	3899	4	US-09-902-540-573	Sequence 573, App
C 19	96.5	4.9	28374	4	US-09-949-016-17508	Sequence 17508, A
C 20	96.5	4.9	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 21	94.5	4.8	4086	4	US-09-614-221A-478	Sequence 478, App
C 22	94.5	4.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 23	94	4.7	1505	4	US-09-291-299A-2	Sequence 2, Appl
C 24	94	4.7	1505	4	US-09-291-299A-2	Sequence 2, Appl
C 25	94	4.7	74881	4	US-09-949-016-15345	Sequence 15345, A
C 26	94	4.7	74914	4	US-09-949-016-12286	Sequence 12286, A
C 27	93.5	4.7	36542	4	US-09-949-016-12149	Sequence 12149, A
C 28	93.5	4.7	36544	4	US-09-949-016-13434	Sequence 13434, A
C 29	92.5	4.7	3816	4	US-09-540-236-1820	Sequence 1820, Ap
C 30	92.5	4.7	6558	4	US-09-491-356C-7	Sequence 7, Appl
C 31	92.5	4.7	63563	4	US-09-596-002-33	Sequence 33, Appl
C 32	92	4.6	77626	4	US-09-949-016-12608	Sequence 1, Appl
C 33	91.5	4.6	2229	4	US-09-214-904-1	Sequence 15698, A
C 34	91.5	4.6	145812	4	US-09-949-016-15698	Sequence 15698, A
C 35	91	4.6	893	4	US-09-270-767-11464	Sequence 11464, A
C 36	91	4.6	5591	4	US-09-949-016-12044	Sequence 12044, A
C 37	91	4.6	5592	4	US-09-949-016-16387	Sequence 16387, A
C 38	91	4.6	10357	3	US-08-961-527-191	Sequence 191, App
C 39	91	4.6	12665	4	US-09-949-016-16388	Sequence 16388, A
C 40	90.5	4.6	1499	4	US-09-509-908-1	Sequence 1, Appl
C 41	90	4.5	20870	4	US-09-949-016-16017	Sequence 16017, A
C 42	90	4.5	35100	2	US-08-770-379-17	Sequence 17, Appl
C 43	90	4.5	35100	3	US-08-757-669A-17	Sequence 17, Appl
C 44	90	4.5	35100	3	US-09-230-371A-17	Sequence 17, Appl
C 45	89.5	4.5	2501	4	US-09-614-891-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```
;          LENGTH: 1100 base pairs
;          TYPE: Nucleic Acid
;          STRANDEDNESS: Single
;          TOPOLOGY: Linear
US-08-978-741-16

Alignment Scores:
Pred. NO.:          2.04e-234          Length:          1100
Score:             1979.00             Matches:          365
Percent Similarity: 100.00%             Conservative:    0
Best Local Similarity: 100.00%           Mismatches:     0
Query Match:       100.00%             Indels:         0
DB:                3                   Gaps:          0

US-09-774-954-2 (1-365) x US-08-978-741-16 (1-1100)

QY      1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB      1 ATGCCCGCGGGCTCTCTGGGACCCGCGCGGTACCTGCTCTACCTGCCCTGCATGGGGCGC 60

QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB      61 TTGGGAACCCAGCCGATCACTTCTGGGCTCTCTGGCATTTGCCAATTTGCCAAGCTGCTAAACCGT 120

QY      41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
DB      121 ACCTTGCTGCTCCCTTCCTGGATGAGTACCGATCAACAGCTCTCTTTCAGCAACCTC 180

QY      61 HisValSerTyrGlnLysTyrPheLysLeuGlnProLeuGlnAlaTyrHisArgValIle 80
DB      181 CATGTGTCTTACCGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240

QY      81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrrProProGluLysArgVal 100
DB      241 AGCTTGAGGATTTATCGAGAAGCTGGCACCCACCTACTGGCCCCCTGGAGACGGGIG 300

QY      101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysLysLysLysLys 120
DB      301 GCATCTGCTTTGAGTGGCAGCCAGCGAGCCGACGATGAAGAACAGCAGTGGCCCATGAAG 360

QY      121 GluGlyAsnProPheGlyProPheTrrAspGlnPheHisValSerPheAsnLysSerGlu 140
DB      361 GAAGAAACCCCTTTGGCCCATTTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

QY      141 LeuPheThrGlyLysSerPheSerAlaSerTyrArgGluGlnTrrSerGlnArgPheSer 160
DB      421 CTTTTACAGGCATTTCTTTCAGTGTCTTCTACAGAGAACAAATGGAGCCAGAGATTTTCT 480

QY      161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
DB      481 CCAAGGAACATCCGTGCTTCCCTTCAGAGAGCCCGAGCCCGAGTTCCTAGAA 540

QY      181 GluHisArgProLeuGlnLysTyrMetValTrrSerAspGluMetValLysLysLysLys 200
DB      541 GAACACAGGCCACTACAGAAGTACATGTATGTGTGACAGCAATGTTGAGNACGGGAGAG 600

QY      201 AlaGlnIleHisAlaHisLeuValArgProTrrValGlyIleHisLeuArgLysSer 220
DB      601 GCCCAGATTATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCACATGGGCTCT 660

QY      221 AspTrrLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
DB      661 GACTTGAAGAACCCTGTGCCATGTGTGAAGACGGGAGCTCGAGGCTCGCATTCATGGCC 720

QY      241 SerProGlnCysValGlyTrrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
DB      721 TCTCGCAGTGTGTGGGCTACAGCCGACGACAGCGGCCCTCCCTACGATGATATGTGC 780

QY      261 LeuProAspLeuLysGlnIleGlnArgAlaValLysLeuTrrValArgSerLeuAspAla 280
DB      781 CTGCCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTGCGTGGATGCC 840

QY      281 GlnSerValTrrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
```

```
DB      841 CAGTCGGTCTAGTTGCTTACTGATTCGAGAGTTAATGCTGAGCTCCAAACAGCTCTTC 900
QY      301 LysGlyLysValLysValSerLeuLysProGluValAlaGlnValAspLeuTrrIle 320
DB      901 AAAGGAAGGTGAAGGTGGTGGAGCTGAAGCTTGAGCTGAGTGGCCAGGTGACCTGTACATC 960
QY      321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
DB      961 CTGGCCCAAGCCGACCACTTATTTGGCACTGTGTCTCTCTCTTCTCCTGCTTGTGAAG 1020
QY      341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
DB      1021 CGGAGCGGGGACCTCCAGCGGAGCGCTCTTCTTCTTCTCGGCATGACAGGCCCTTAAG 1080
QY      361 LeuArgAspGluPhe 365
DB      1081 CTGGGACGAGTTC 1095

RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 16
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-16

Alignment Scores:
Pred. No.:          2.04e-234          Length:          1100
Score:             1979.00             Matches:          365
Percent Similarity: 100.00%             Conservative:    0
Best Local Similarity: 100.00%           Mismatches:     0
Query Match:       100.00%             Indels:         0
DB:                3                   Gaps:          0

US-09-774-954-2 (1-365) x US-09-333-729A-16 (1-1100)

QY      1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB      1 ATGCCCGCGGGCTCTCTGGGACCCGCGCGGTACCTGCTCTACCTGCCCTGCATGGGGCGC 60

QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB      61 TTGGGAACCCAGCCGATCACTTCTGGGCTCTCTGGCATTTGCCAATTTGCCAAGCTGCTAAACCGT 120

QY      41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
DB      121 ACCTTGCTGCTCCCTTCCTGGATGAGTACCGATCAACAGCTCTCTTTCAGCAACCTC 180

QY      61 HisValSerTyrGlnLysTyrPheLysLeuGlnProLeuGlnAlaTyrHisArgValIle 80
DB      181 CATGTGTCTTACCGAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240

QY      81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrrProProGluLysArgVal 100
DB      241 AGCTTGAGGATTTATCGAGAAGCTGGCACCCACCTACTGGCCCCCTGGAGACGGGIG 300

QY      101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysLysLysLysLys 120
DB      301 GCATCTGCTTTGAGTGGCAGCCAGCGAGCCGACGATGAAGAACAGCAGTGGCCCATGAAG 360
```

```
Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 361 GAAGAAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 421 CTTTTTACAGGCATTTCTTCAGTGCTTCTCAGAGAAACAATGGAGCCAGAGATTTCT 480
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db 481 CCAAGGAACATCCGGTGCTTGGCCCTGCCAGAGCCCAAGCCAGTTTCCCGCTCTAGAA 540
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 541 GAACACAGGCACACTACAGAGTACATGGTATGTCAGACGAAATGGTGAAGACGGAGAG 600
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 601 GCCCAGATTATGCCACCTTGTCCGCCCTATGTGGGCATTCATCTGCGCATTTGGCTCT 660
Qy 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 661 GACTGGAAGAACCCCTGTGCCATCTGTAAGACGGGACTCGAGGCTCGCACTTCATGGCC 720
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 721 TCCTCCGAGTGTGTGGCTACGCCGCTACGCCGAGCAGCGGCCCTCCAGCATGATGTGTC 780
Qy 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db 781 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCTGCTGGATGCC 840
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 841 CAGTCGCTTACGTTGTCTACTGATTCGAGAGTTATGTGCTGAGCTCCCAACAGCTCTTC 900
Qy 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 901 AAAGGAAGGTGAAGTGTGAGCTGAGCCTGAGCTGAGGTGGCCCAAGTCTGATCATC 960
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 961 CTGCGCAAGCCAGCCACTTTATTTGGCACTGTGCTCTCTCTCTCTCTCTCTCTCTCT 1020
Qy 341 ArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArgProProLys 360
Db 1021 CCGGAGCGGACCTCCAGGGGAGCGCTCTTTCTTCGGCATGGACAGGCCCTTAAG 1080
Qy 361 LeuArgAspGluPhe 365
Db 1081 CTGCGGACGAGTTTC 1095
```

RESULT 3

```
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1300 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-978-741-4

Alignment Scores:
Pred. No.: 2,69e-234 Length: 1300
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-774-954-2 (1-365) x US-08-978-741-4 (1-1300)
Qy 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 136 ATCCCCCGGGCTCTTGGGACCCCGCGGTACTGCTCTACTGCCCTGCATGGGGCGC 195
Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 196 TTTTGGAAACAGGCCGATCACTTCTTGGGCTCTCTGGCAATTGCAAGCTGCTAAACCGT 255
Qy 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProPheThrAsnLeu 60
Db 256 ACCTTGGCTGTCTCTCTTGGATTGAGTACCAAGCTCACAAGCTCTCTTCCACCACTC 315
Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db 316 CATGTGCTTACCAAGATCTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGTCATC 375
Qy 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
Db 376 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCTGGCCCTGAGAGCGGGTG 435
Qy 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db 436 GCATCTGCTTTGAGGTGGCAGCCAGCGAAGCCCAAGATAGAAGACGTCGCCCATGAAG 495
Qy 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db 496 GAAGGAACCCCTTTGGCCCATCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db 556 CTTTTTACAGGCATTTCTTCAGTGCTTCTCAGAGAAACAATGGAGCCAGAGATTTCT 615
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db 616 CCAAGGAACATCCGGTGCTTGGCCCTGCCAGAGCCCAAGCCAGTTTCCCGCTCTAGAG 675
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db 676 GAACACAGGCACACTACAGAGTACATGGTATGTCAGACGAAATGGTGAAGACGGAGAG 735
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
```

Db 736 GCCAGATTATGCCACCTTTGTCGGGCCCTATGTGTGGGCAATTCATCTGGCAATGGCTCT 795
Qy 221 AsptTpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 796 GACTGGAAAGACGGCTGTGGCATGCTGAAGAGACGGGACTGCGAGGCTCGCACTTCATGGCC 855
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 856 TCTCCGCAGTGTGTGGCTACAGCCGACAGCCGAGGCCCCCTCCACGATGACTATGTGC 915
Qy 261 LeuProAspLeuLysGluLeGlnArgAlaValLysLeuTyrValAlaArgSerLeuAspAla 280
Db 916 CTGCTTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTGGCTGGATGCC 975
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 976 CAGTCGGTCTAGTGTGCTACTGATTCGAGAGTTATGTGCCCTGAGCTCCAAACAGCTCTTC 1035
Qy 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 1036 AAAGGAGAGTGAAGTGGTGAGCCTGAAGCCTGAGGTGGGCCAGGTCGACCTGTACATC 1095
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 1096 CTCGGCCAGCGGACCACTTATTGGCACTGTGCTCTCTCTTCACTGCTTGTGAAG 1155
Qy 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProlys 360
Db 1156 CGGAGCGGAGCTCCAGGAGGCGCTCTTCTTTCGGCATGGACAGGCCCCCTAAG 1215
Qy 361 LeuArgAspGluPhe 365
Db 1216 CTGCGGACGAGTTTC 1230

RESULT 4

US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: Spellman, Michael W.
; FILE REFERENCE: P1041PIDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333, 729A
; PRIOR FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Alignment Scores:
Pred. No.: 2,69e-234 Length: 1300
Score: 1979.00 Matches: 365
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 3
DB:

US-09-774-954-2 (1-365) x US-09-333-729A-6 (1-1300)

Qy 1 MetProAlaGlySerTyrAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 136 ATGCCCGCGGGCTCTCTGGGACCCGCGGCTTACTGCTCTACTGCTCTGATGGGGCGC 195
Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 196 TTGGGAAACAGCGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGAGCTGCTAAACCGT 255

RESULT 5

US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 610076

Qy 41 ThrLeuAlaValProProTrrPleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db 256 ACCTTGGCTGTCCCTTCTTGGATTGAGTACCAAGATCATCAAGCCCTCTCTTTCACCAACCTC 315
Qy 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db 316 CATGTGCTCTACCAAGATGACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 375
Qy 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrrProProGluLysArgVal 100
Db 376 AGCTTGGAGGATTTTCATGGAGAGCTGGCACCCACCACTGGCCCTCCCTGAGAGCGGGTG 435
Qy 101 AlaTyrCysPheGluValAlaAlaGlnAtqSerProAspLysLysThrCysProMetLys 120
Db 436 GCATATGCTTTTGAAGTGGGAGCCAGCGAGGCCAGATAAGAGACGTGGCCCATGAAG 495
Qy 121 GluGlyAsnProPheGlyProPheTrrPhePheHisValSerPheAsnLysSerGlu 140
Db 496 GAAGGAAACCCCTTTGGGCCCATCTGGGATCAGTTTCATGTGAGTTTCACAGTCCGAG 555
Qy 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrrPheSerGlnArgPheSer 160
Db 556 CTTTTTACAGGCATTTCTTTCAGTGTCTTCTCAGAGAAACAATGGAGCCAGAGATTTTCT 615
Qy 161 ProLysGluHisProValLeuAlaLeuProGlyValaProAlaGlnPheProValLeuGlu 180
Db 616 CCAAGGAACATCCGGTGTCTTGGCTGGCAGGAGGCCAGCCAGTTCCTCCGTCCTAGAG 675
Qy 181 GluHisArgProLeuGlnLysTyrMetValTrrPheSerAspGluMetValLysThrGlyGlu 200
Db 676 GAACAGAGCCACTACAGAGTACATGTGTATGTGTGAGCAAAATGTGAAGACGGAGAG 735
Qy 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db 736 GCCAGATTTCATGCCCACTTGTTCGGCCCTATGTGGGCAATTCATCTCGCATTTGGCTCT 795
Qy 221 AsptTrrLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db 796 GACTTGAAGAACGCTGTGCCATGCTGAAGAGCGGACTGCGAGGCTCGCACTTCATGGCC 855
Qy 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db 856 TCTCCGCACTGTGTGGGCTACAGCCGAGCAGCAGCGGCCCTCCATGACATGATGTGC 915
Qy 261 LeuProAspLeuLysGluLeGlnArgAlaValLysLeuTrrValArgSerLeuAspAla 280
Db 916 CTGCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGGTCGCTGATGCC 975
Qy 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
Db 976 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCAGCTGACCTGTACATC 1035
Qy 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db 1036 AAAGGAGAGTGAAGTGGTGAGCCTGAAGCCTGAGGTGGGCCAGGTCGACCTGTACATC 1095
Qy 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db 1096 CTCGGCCAGCGGACCACTTATTGGCACTGTGCTCTCTCTTCACTGCTTGTGAAG 1155
Qy 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProlys 360
Db 1156 CGGAGCGGAGCCTCCAGGAGGCGCTTCTTCTTTCGGCATGGACAGGCCCCCTAAG 1215
Qy 361 LeuArgAspGluPhe 365
Db 1216 CTGCGGACGAGTTTC 1230

; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,741
 ; FILING DATE: 26-No. 6100076-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/792498
 ; FILING DATE: 31
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1041P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1514 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear
 ; US-08-978-741-1

Alignment Scores:
 Pred. No.: 3,47e-234 Length: 1514
 Score: 1979.00 Matches: 365
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-09-774-954-2 (1-365) x US-08-978-741-1 (1-1514)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 DB 1 ATGCCCGGGGCTCTGGGACCCGGCGGTACCTGCTCTACTGCCCTGCGATGGGCGC 60
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 DB 61 TTTGGGAACACGGCCGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
 QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
 DB 121 ACCTTGGCTGTCCCTCTGGATTGAGTACCAGCATCACAGCCCTCCCTTACCAACCTC 180
 QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
 DB 181 CATGTGCTTACCAAGAAGTACTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATC 240
 QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
 DB 241 AGCTTGGAGGATTCATGGAGAAGCTGGACCCACCTGCCCCCTGGAGAAGCGGGT 300
 QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 DB 301 GCATACTGCTTTGAGTGGGAGCCAGCCAGAGCCAGATAGAGAGCGTGGCCCATGAAG 360
 QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140

DB 361 GAAGGAACCCCTTTGGCCCATTTCTGGGATCAGTTTTCATGTGAGTTTCAACAGTCCGAG 420
 QY 141 LeuPheThrGlyLysSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
 DB 421 CTTTTTACAGGCATTTCCCTTCAGTGTCTCTACAGAGAACAATGGAGCCAGAGATTTCT 480
 QY 161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
 DB 481 CCBAAGGAACATCCGGTGTCTTGGCCCTGCCAGGAGCCAGCCAGCTTCCCGCTCTAGAA 540
 QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 DB 541 GAACACAGGCCACTACAGAAGTACATGGTATGGTACAGCAATGGTGAAGCGGAGAG 600
 QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 DB 601 GCCCAGATTTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCCATGGCTCT 660
 QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 DB 661 GACTGGAAGACCGCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720
 QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
 DB 721 TCTCCGCGAGTGTGTGGGCTACAGCCGACGACAGCGGCCCTCCATGATGATATGTGC 780
 QY 261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
 DB 781 CTGCCCTGACCTGAAGGAGATCCAGAGGCTGTGAGCTCTGGGTGAGGTCGCTGGATGCC 840
 QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
 DB 841 CAGTCGGTCTACCTTGTCTACTGATTCGAGAGTTATGTGCTTCAGGCTCCCAACAGCTCTC 900
 QY 301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
 DB 901 AAAGGGAAGGTGAAGGTGTGAGCTGAAGCTGAGGTGGCCCGGAGGTGCGACCTGTATCATC 960
 QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
 DB 961 CTCGGCCAAAGCCACCACTTTATTTGGCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
 QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
 DB 1021 CGGAGCGGGACCTCCAGGGGAGCGGCTCTCTTCTTCTGCGCATGGAGAGGCCCTTAAG 1080
 QY 361 LeuArgAspGluPhe 365
 DB 1081 CTGGGAGCAGATTC 1095
 RESULT 6
 US-09-333-729A-2
 ; Sequence 2, Application US/09333729A
 ; Patent No. 6270987
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Yang
 ; APPLICANT: Spellman, Michael W.
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; FILE REFERENCE: P1041P1-Substitute
 ; CURRENT APPLICATION NUMBER: US/09/333,729A
 ; CURRENT FILING DATE: 1999-06-15
 ; PRIOR APPLICATION NUMBER: US 08/798,741
 ; PRIOR FILING DATE: 1997-11-26
 ; NUMBER OF SEQ ID NOS: 21
 ; SEQ ID NO 2
 ; LENGTH: 1514
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-09-333-729A-2
 Alignment Scores:
 Pred. No.: 3,47e-234 Length: 1514

Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-774-954-2 (1-365) x US-09-333-729A-2 (1-1514)

QY	1	MetProAlaGlySerTyrAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	1	ATGCCCCGGGCTCTGGGACCGCGCGGTACTCTGCTCTACTGCGGCTGATGGGGGCG	60
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	61	TTTGGGACACAGCGCGATCATCTCTTGGGCTCTCTGGCATTTGGCAAGCTGCTTAACCGT	120
QY	41	ThrLeuAlaValProProTrrPilleGlyTyrGlnHisLysLysProProPheThrAsnLeu	60
DB	121	ACCTTGCGTGTCTCCCTTGTGATGTGAGTACAGCATCAAGACCTCTCTTACCAACCTC	180
QY	61	HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaLysHisArgValle	80
DB	181	CATGTGTCTTACAGAAGTACTTCAAGCTGGAGCGGCTCCAGGCTTACCATCGGCTCATC	240
QY	81	SerLeuGlnAlaPheMetGluLysLeuAlaProThrHisTyrProGluLysArgVal	100
DB	241	AGCTTGGAGGATTCATGGAGAAGCTGGCACCCACCACTGGCCCCCTGAGAGGGGGTG	300
QY	101	AlaTyrCysPheGluValAlaAlaGlnAlaSerProAspLysLysThrCysProMetLys	120
DB	301	GCATCTGCTTTGAGGTGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG	360
QY	121	GluGlyAsnProPheGlyProPheTrrPAspGlnPheHisValSerPheAsnLysSerGlu	140
DB	361	GAAGGAACCCCTTTGGGCCCATCTGGGATCAGTTTCAATGAGTTTCAACAAAGTCGGAG	420
QY	141	LeuPheThrGlyLysPheSerAlaSerTyrArgGluGlnTrrPserGlnArgPheSer	160
DB	421	CTTTTACAGGCATTTCTCTCAGTGCTTCTCAGAGAACCAATGGAGCCAGAGATTTCT	480
QY	161	ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu	180
DB	481	CCAAAGGAACATCCGGTCTTGGCTGCCAGGAGGCCAGGCCAGGCTTCCCGCTCTAGAA	540
QY	181	GluHisArgProLeuGlnLysTyrMetValTrrPserAspGluMetValLysThrGlyGlu	200
DB	541	GAACAGGCCACTACAGAGTACATGGTATGGTTCAGAGAAATGCTGAAGACGGGAGAG	600
QY	201	AlaGlnIleHisAlaHisLeuValArgProTrrValGlyLleHisLeuArglleGlySer	220
DB	601	GCCAGATTCATGCCCATCTTGTCCGGCCCTATGTGGGCATTCATCTGGCATTGGCTCT	660
QY	221	AspTrrLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla	240
DB	661	GACTGGAAGAACGCTGTGTGCCATCTCTGAAGGACGGGACTGCGAGGCTCGCATGGCC	720
QY	241	SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys	260
DB	721	TCTCGCAGTGTGTGGCTACAGCGCGCAGCACAGCGGCCCCCTCACGATGACTATGTC	780
QY	261	LeuProAspLeuLysGluIleGlnAlaValLysLeuTrrValArgSerLeuAspAla	280
DB	781	CTGCTGACTGAGAGATCAAGAGGCTGTGAGAGCTCTGGGTGAGTCCGTGATGCC	840
QY	281	GlnSerValTrrValAlaThrAspSerGluSerTrrValProGluLeuGlnGlnLeuPhe	300
DB	841	CAGTGGGTCTAGTGTGCTACTGATTCGAGAGTATGTGGCTTGAGCTCCAAACAGCTTTC	900
QY	301	LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTrrLys	320
DB	901	AAAGGAAGGTGAAGGTGTGAGCTTGAAGCTTGAAGGTGGGCGGCGGCTGATCATC	960
QY	321	LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys	340

US-09-774-954-2 (1-365) x US-08-978-741-5 (1-11284)

QY	1	MetProAlaGlySerTrrAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg	20
DB	4236	ATGCCCCGGGCTCTGGGACCGCGCGGTACTCTCTACTGCCCCCTGATGGGGCGC	4295
QY	21	PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg	40
DB	4296	TTTGGGAACAGGCGGATCATCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT	4355
QY	41	ThrLeuAlaValProProTrrPilleGlyTyrGlnHisLysLysProProPheThrAsnLeu	60

US-08-978-741-5

Sequence 5, Application US/08978741

Patent No. 6100076

GENERAL INFORMATION:

APPLICANT: Yang Wang, Michael W. Spellman

TITLE OF INVENTION: O-Fucosyltransferase

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSES: Genentech, Inc.

STREET: 1 DNA Way

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA: US/08/978,741

APPLICATION NUMBER: 26-No. 6100076-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/792498

FILING DATE: 31

ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.

REGISTRATION NUMBER: 39,044

REFERENCE/DOCKET NUMBER: P1041P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-1489

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 11284 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-978-741-5

Alignment Scores:

Pred. No.: 9,61e-233 Length: 11284

Score: 1979.00 Matches: 365

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 3 Gaps: 0

Db 4356 ACCTGGCTGCTCCCTCTGGATTGAGTACAGCATCAAGCGCTCCTTTCCACCAACCTC 4415
 QY 61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
 Db 4416 CATGTGCTCTACAGAGTACTTCAAGCTGGAGCCCTCCAGCTTACCATCGGGTCATC 4475
 QY 81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgVal 100
 Db 4476 AGCTTGGAGATTTCATGAGAGCTGGACCCACCTCTGGCCCTCGAAGCGGGTG 4535
 QY 101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
 Db 4536 GCATACCTGCTTGGAGTGGAGCCAGGAGCCAGATAGAGAGAGCTGCCCATGAAG 4595
 QY 121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
 Db 4596 GAAGGAACCCCTTGGCCCATTTCTGGGATTCAGTTTCATGTGATTTCAACAGTCGGAG 4655
 QY 141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
 Db 4656 CTTTTTACAGGCAATTCCTTCAGTGCTTCTCAGAGAACATGGAGCCAGAGATTTCT 4715
 QY 161 ProLysGluHisProValLeuAlaLeuProGlyValaProAlaGlnPheProValLeuGlu 180
 Db 4716 CCAAGGACATCGGTGCTTGGCTGCCAGGAGCCAGCCAGTTCCTCGCTTAGAG 4775
 QY 181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
 Db 4776 GAACACAGGCCACTACAGAGTACATGATGTGTGACAGCAATGGTGAAGCGGAGAG 4835
 QY 201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
 Db 4836 GCCCAGATTCATGCCACCTTGTCCGGCCCTATGTGGGCATTCATTCGCCATGCTCT 4895
 QY 221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
 Db 4896 GATCGAGAGACGCTGTGCTGCTGAGAGCGGAGCTCGAGCTCGCACTTCATGGCC 4955
 QY 241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
 Db 4956 TCTCCGAGTGTGGGCTACAGCGCAGCACAGCGGCCCTCCAGTACGATGATGTGC 5015
 QY 261 LeuProAspLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
 Db 5016 CTCCTGACCTGAAGAGATCCAGAGGCTGTGAAGCTCTGGGTGAGTCCGCTGAGTGC 5075
 QY 281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnLeuPhe 300
 Db 5076 CAGTCGGTCTACGTTGCTACTGATTCGAGAGTTATGTGCTGAGCTCCACAGCTCTTC 5135
 QY 301 LysGlyLysValLysValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
 Db 5136 AAAGGAAGGTGAAGTGTGTGAGCTGAAGCTGAGGTGGCCAGGTCGACCTGTATCATC 5195
 QY 321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerPheThrAlaPheValLys 340
 Db 5196 CTCGGCCAAAGCCGACCTTATTTGGCACTGTGCTCTCTTCCTTCATCTGCTTTGAG 5255
 QY 341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProLys 360
 Db 5256 CGGAGCGGGACCTTCAGGGGAGGCGCTCTCTCTTCCTTCGCGATGAGAGCCCTTAAG 5315
 QY 361 LeuArgAspGluPhe 365
 Db 5316 CTGCGGAGCGAGTTC 5330

RESULT 8

US-08-978-741-7
 ; Sequence 7, Application US/08978741
 ; Patent No. 6100076
 ; GENERAL INFORMATION:
 ; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase

NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 1 DNA Way
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WinPatIn (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,741
 FILING DATE: 26-No. 6100076-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/792498
 FILING DATE: 31
 ATTORNEY/AGENT INFORMATION:
 NAME: Svoboda, Craig G.
 REGISTRATION NUMBER: 39,044
 REFERENCE/DOCKET NUMBER: P1041PI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650/225-1489
 TELEFAX: 650/952-9881
 INFORMATION FOR SEQ ID NO: 7:
 LENGTH: 5009 base pairs
 TYPE: Nucleic Acid
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 US-08-978-741-7

Alignment Scores:
 Pred. No.: 1,24e-216 Length: 5009
 Score: 1844.00 Matches: 343
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 93.18% Indels: 0
 DB: Gaps: 0

US-09-774-954-2 (1-365) x US-08-978-741-7 (1-5009)

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuAsnArgThrLeu 42
 Db 2 AACACGGCCGATCACCTTCTGGGCTCTTGGCAATTTGCAAGCTGCTTAAACCCGATCCTTG 61
 QY 43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
 Db 62 GCTGTCCCTCTTGGATTGATGATACAGCATCACAGCCCTCTTCCACCAACCTCCATGTG 121
 QY 63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
 Db 122 TCCTACCAAGTACTTCAAGCTGGAGCCCTCCAGGCTTACCATCGGGTCATCAGCTTG 181
 QY 83 GluAspPheMetGluLysLeuAlaProThrHisTrpProGluLysArgValAlaTyr 102
 Db 182 GAGGATTTCAAGAGAGCTGGACCCACCTCCAGGCTTACCATCGGGTCATCAGCTTG 241
 QY 103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
 Db 242 TGCTTTGAGTGGCAGCCAGCCAGAGCCAGATAGAGAGCTGCCCATGAGGAGGA 301
 QY 123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
 Db 302 AACCCCTTGGCCCATTTCTGGGATTCAGTTTCATGTGATTTCAACAGTCGAGCTTTT 361
 QY 143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
 Db 362 ACAGGCATTTCTTTCAGTGTCTTCTCAGAGAACAAATGGAGCCAGAGATTTTCTCCAAG 421

Qy	163	GlutHisProValLeuAlaLeuProGluYalaProAlaGlnPheProValLeuGluLys	182
Db	422	GAACATCCGGTGTCTGCCCTCCAGGAGCCCGAGCCAGTGTCCCGTCTCAGAGAACAC	481
Qy	183	ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln	202
Db	482	AGGCCACTCAGAGAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGCCCGAC	541
Qy	203	IleHisAlaHisLeuValArgProTyrValGlyVilHisLeuArgIleGlySerAspTrp	222
Db	542	ATTTCATGCCCACTTGTCCGGCCCTATGTGGGCAVTCATCTGCACATTTGGCTCTGACTGG	601
Qy	223	LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro	242
Db	602	AAGAACGCTGTGCCATGCTGAAGACGGGACTGCAGGCTCGCACTTCATGGCCCTCTCG	661
Qy	243	GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro	262
Db	662	CAGTGTGTGGGCTACAGCGCAGCACAGCGGCCCTCCATGATGACTATGTGCTTGCT	721
Qy	263	AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer	282
Db	722	GACCTGAAGGAGATCCAGAGGGCTGTGAGAGCTCTGGGTGAGTGCCTGAGATGCCAGTGC	781
Qy	283	ValTyrValAlaThrAspSerGlnSerTyrValProGluLeuGlnLeuPheLysGly	302
Db	782	GTCTACGCTGTCTACTGATTCCGAGAGTTATGTGTGCCCTGAGCTCCAACAGCTCTTCAAAGGG	841
Qy	303	LysValLysValIValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly	322
Db	842	AAGCTGAAGGTGTGAGGCTGAAGCCTGAGGTGGCCAGGCTCGACCTGTACATCTCTGGC	901
Qy	323	GlnAlaAspHisPheIleGlyValAsnCysValSerSerPheThrAlaPheValLysArgGlu	342
Db	902	CAAGCGGACCCACTTTATGGCAACTGTGTCTCTCTCTTCACTGCCTTTGTGAAGCGGGAG	961
Qy	343	ArgAspLeuGlnClyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg	362
Db	962	CGGACCTCCAGGGAGCGCGTCTCTCTTTCGGGATGGACAGGCCCTCTAAGCTGGG	1021
Qy	363	AspGluPhe	365
Db	1022	GACGAGTTC	1030
RESULT 9			
US-09-333-729A-8			
; Sequence 8, Application US/09333729A			
; Patent No. 6270987			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Yang			
; APPLICANT: Spellman, Michael W.			
; TITLE OF INVENTION: O-Fucosyltransferase			
; FILE REFERENCE: P1041PId1-Substitute			
; CURRENT APPLICATION NUMBER: US/09/333, 729A			
; CURRENT FILING DATE: 1999-06-15			
; PRIOR APPLICATION NUMBER: US 08/798, 741			
; PRIOR FILING DATE: 1997-11-26			
; NUMBER OF SEQ ID NOS: 21			
; SEQ ID NO 8			
; LENGTH: 5009			
; TYPE: DNA			
; ORGANISM: Homo Sapien			
US-09-333-729A-8			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
Length:			
Matches:			
Conservative:			
Mismatch:			
Indels:			
Gaps:			
5009			
343			
100.00%			
100.00%			
93.18%			
3			

US-09-774-954-2 (1-365) x US-09-333-729A-8 (1-5009)


```

RESULT 10
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14353

Alignment Scores:
Pred. No.: 4,55e-93 Length: 1320
Score: 837.00 Matches: 168
Percent Similarity: 61.50% Conservative: 62
Best Local Similarity: 44.92% Mismatches: 118
Query Match: 42.29% Indels: 26
DB: 9 Gaps: 4

US-09-774-954-2 (1-365) x US-09-270-767-14353 (1-1320)
QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
DB 75 GATCCCAATGGCTACTCTACCTACCTACTGCTGATGGAGCGCTTTGGCAACACGAGCGGAC 134
QY 27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
DB 135 CACTTCCTGGATCAATGGCTTTCGCGCAAGCGCTTAATCGCACCTGTATCTCGCGCGG 194
QY 47 TrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLys 66
DB 195 TGGGTGGAGTATCGT-----AGGGTGAACTCGGATCCCGTCAGGTACCGTTCAACACA 248
QY 67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMet 86
DB 249 TATTTGAAGTGAGCGCCCTGAAGGAATACCATCGGTCATCACCATTGGCAGATTTCATG 308
QY 87 GluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu--- 105
DB 309 TGGCACCTGGCCGACGACATTTGGCCAGATCGGAGCGAGTGTCAATTTTCTACAAAGAA 368
QY 106 -----ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
DB 369 CGATATAGCTTCAGCAGGAGAGAAGACGATCCAGACAAGCCCAATTGCCACGCCAAGGAT 428
QY 122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
DB 429 GGCAATCTTTGGTCCCTTTGGACACATTTTCACATACACTTTGTGCGGTGAGGTTTC 488
QY 142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGln-----TrpSerGlnArg 158
DB 489 TATGGCGCACTTCATTTTGATGTGATCATAGCAAGCGAGTCCCAAGTGGCAGACCAAA 548
QY 159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
DB 549 TATCTCTCGAATCATATCCCGTACTCGCGTTCACCGGAGCTCGGCTAGTTTTCCTGTT 608
QY 179 LeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr 198
DB 609 CAGTAGAGAACTCAGAGCTGCAGCGTACTTGCAGTGGAGTCAACGGTATAGGAGCA 668
QY 199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
DB 669 TCTAAGGATTTTCATCCGAGAGCAGTGTGCTTGGGGTGCCTTTTGGGCAATTCATCTCGC 728
QY 218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237

```

```

DB 729 AACGGTATCGATTGGGTGAGAGCCCTGTGAGCAGCTCAAGGAT-----AGCCAGCAT 779
QY 238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
DB 780 CTGTTTCCCTCCCGCAGTGTCTTGGGCTATAAATAAGAACGTGGTGCA---CTCTACCCG 836
QY 258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
DB 837 GAGCTCTGATCCCTCCCAAGGAGCGCATCATCCGCCAGCTAAAGAGAACCATTAAGAAC 896
QY 278 Leu-----AspAlaGlnSerValTyrValAlaThrAspSerGlu 290
DB 897 GTGCGCAAACTCAGCCGACCAACAAATCAATCAGTTTTCGTGGCGTCAGACTCCAAT 956
QY 291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysValSerLeu 309
DB 957 CACATGATTGGTGAACATAACACGGCCCTTAGTCGCATCGGCATCAGTGTGCACAAGCTG 1016
QY 310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
DB 1017 ACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGC 1076
QY 330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
DB 1077 AACTGTATATCTCTTACTCGGCATTTCGTAAAAAGGGAACGAGATGTGCACGGTTTCCA 1136
QY 350 SerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
DB 1137 TCGTACTTCTGGGATTC-----CCCAAGGAAAAAGGAT 1169

RESULT 11
US-09-270-767-30507
; Sequence 30507, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30507
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-30507

Alignment Scores:
Pred. No.: 2,05e-32 Length: 722
Score: 343.00 Matches: 80
Percent Similarity: 57.87% Conservative: 34
Best Local Similarity: 40.61% Mismatches: 67
Query Match: 17.33% Indels: 16
DB: 6 Gaps: 4

US-09-774-954-2 (1-365) x US-09-270-767-30507 (1-722)
QY 176 PheProValLeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMet 195
DB 2 TTTCTCTGTTTCAGTAGAAGAACTGCAAGCTGCAGCGCTACTTTCAGTGGAGTCAACGGTAT 61
QY 196 ValLysThrGlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIle 214
DB 62 AGGGAAGCATCTAAGGATTTTCATCCGAGAGCAGTGTGCTCGGGGTGCCTTTTGGGGCAT 121
QY 215 HisLeuArgIleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAla 234
DB 122 CATCTGCGCAACGGTATCGATTGGTGAGAGCTTGTGAGCAGCTGTCAGCAGCTCAAGGAT 172
QY 235 GlySerHisPheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaPro 254

```

```
Db 173 AGCCAGCATCTCTTTGGCTCGCCGCAAGTGGTGGGCTATATAAATGAACGTGTGCA--- 229
Qy 255 LeuThrMetThrMetCysLeuProAspLeuLysGluileGlnArgAlaValLysLeuTrp 274
Db 230 CTATCCCGAGCTCTGTCATCCCTCAAGAGGGGGGATCATCCGCCAGCTAAAGAGAAC 289
Qy 275 ValArgSerLeu-----AspAlaGlnSerValTyValAlaThr 287
Db 290 ATTAAGACGTGGCCAACTCAGCCGGAACAAGAAATCAAATCAGTTTCGTGGCGTCA 349
Qy 288 AspSerGluSerTyValProGluLeuGln---GlnLeuPheLysGlyLysVal 306
Db 350 GACTCCATCATCATGATTGGTGAACCTAAACACGGCCCTTAGTCGCATCGGCATCATGTG 409
Qy 307 ValSerLeuLysProGluValAlaGlnValAspLeuTyTlleLeuGlyGlnAlaAspHis 326
Db 410 CACAAGCTCAGCGAGGATGATCTTACTCGGACTTTCGGACAGTCCGAACAC 469
Qy 327 PheileGlyAsnCysValSerPheThrAlaPheValLysArgGluArgAspLeuGln 346
Db 470 TTATCGGCAACTGTATATCTCTTACTCGGCATTCGTAAGAAAGGAAACGAGATGTGCAC 529
Qy 347 GlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
Db 530 GGTTTTCATCTACTTCTGGGATTC-----CCCAAGGAAAGAGT 571

RESULT 12
US-09-270-767-28905/c
; Sequence 28905, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28905
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-28905

Alignment Scores:
Pred. No.: 9,08e-13 Length: 998
Score: 186.00 Matches: 41
Percent Similarity: 67.44% Conservative: 17
Best Local Similarity: 47.67% Mismatches: 24
Query Match: 9.40% Indels: 4
DB: 4 Gaps: 2

US-09-774-954-2 (1-365) x US-09-270-767-28905 (1-998)
Qy 279 AspAlaGlnSerValTyValAlaThrAspSerGluSerTyValProGluLeuGln--- 297
Db 975 GAAATCAAAATCAGTTTCGTGGCTCAGACTCCATCATCATGATGGTGAACATAACACG 916
Qy 298 GlnLeuPheLysGlyLysValValSerLeuLysProGluValAlaGlnValAsp 317
Db 915 GCCCTTAGTCGATCGGCATCAGTGTGCACAAGCTCAGCGAGGATGATCCTTACCTGGAC 856
Qy 318 LeuTyTlleLeuGlyGlnAlaAspHisPheileGlyAsnCysValSerPheThrAla 337
Db 855 TTGGCCATCTCGGACAGTCGAACCACTTTATCGGCAACTGTATATCTCTTACTCGGCA 796
Qy 338 PheValLysArgGluArgAspLeuGlnGlyArgProSerPhePheGlyMetAspArg 357
Db 795 TTGTAAGAAGGACGAGATGTGCACGGTTTCCATCTGATCTCTGGGATTC----- 742
Qy 358 ProProLysLeuArgAsp 363
Db 741 ---CCCAAGGAAAGAGT 727
```

```
RESULT 13
US-09-270-767-13022/c
; Sequence 13022, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13022
; LENGTH: 2034
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-13022

Alignment Scores:
Pred. No.: 2,95e-12 Length: 2034
Score: 186.00 Matches: 41
Percent Similarity: 67.44% Conservative: 17
Best Local Similarity: 47.67% Mismatches: 24
Query Match: 9.40% Indels: 4
DB: 4 Gaps: 2

US-09-774-954-2 (1-365) x US-09-270-767-13022 (1-2034)
Qy 279 AspAlaGlnSerValTyValAlaThrAspSerGluSerTyValProGluLeuGln--- 297
Db 2011 GAAATCAAAATCAGTTTCGTGGCTCAGACTCCATCATCATGATGGTGAACATAACACG 1952
Qy 298 GlnLeuPheLysGlyLysValValSerLeuLysProGluValAlaGlnValAsp 317
Db 1951 GCCCTTAGTCGATCGGCATCAGTGTGCACAAGCTCAGCGAGGATGATCCTTACCTGGAC 1892
Qy 318 LeuTyTlleLeuGlyGlnAlaAspHisPheileGlyAsnCysValSerPheThrAla 337
Db 1891 TTGGCCATCTCGGACAGTCGAACCACTTTATCGGCAACTGTATATCTCTTACTCGGCA 1832
Qy 338 PheValLysArgGluArgAspLeuGlnGlyArgProSerPhePheGlyMetAspArg 357
Db 1831 TTGTAAGAAGGACGAGATGTGCACGGTTTCCATCTGATCTCTGGGATTC----- 1778
Qy 358 ProProLysLeuArgAsp 363
Db 1777 ---CCCAAGGAAAGAGT 1763

RESULT 14
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12127
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12127
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 35.1971 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGLYLXPXMGK.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot:.*
2: uniprot_trembl:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	320	98.2	61	1	OPUL CRIGR
2	305	93.6	393	1	OPUL_MOUSE
3	305	93.6	393	2	Q8CGR4
4	302	92.6	395	2	Q6EV70
5	295	90.5	388	1	OPUL_HUMAN
6	295	90.5	388	1	OPUL_PANTR
7	288	88.3	351	2	Q7YRE6
8	288	88.3	391	2	Q7YRE7
9	286	87.7	378	2	Q7YRZ3
10	273	83.7	391	2	Q6EV66
11	272	83.4	380	2	Q8AWB4
12	250	76.7	380	2	Q640S0
13	250	76.7	396	2	Q6EV71
14	249	76.4	395	2	Q7T0Z8
15	247	75.8	384	2	Q70AG7
16	247	75.8	387	2	Q6EV65
17	234	71.8	380	2	Q8AXS8
18	177	54.3	402	1	OPUL DROME
19	177	54.3	402	2	Q86SR7
20	177	54.3	402	2	Q659S1
21	171	52.5	399	2	Q7QH57
22	170	52.1	381	2	Q6EV67
23	164	50.3	402	2	Q6EV68
24	156	47.9	385	2	Q659P9
25	150.5	46.2	353	2	Q659S0
26	150	46.0	364	2	Q8MSR1
27	142	43.6	381	1	OPUL_CABEL
28	136	41.7	434	2	Q659S2
29	77.5	23.8	602	2	Q9L0H6
30	72.5	22.2	601	2	Q82DS5
31	68.5	21.0	569	2	Q9CCV1

32	68.5	21.0	585	2	Q59530
33	68	20.9	375	2	Q76B90
34	67.5	20.7	354	1	OPSI_ASTFA
35	67.5	20.7	376	1	FUT1_RAT
36	67.5	20.7	538	2	Q7XC59
37	67.5	20.7	538	2	Q9AUT5
38	66.5	20.4	376	1	FUT1_MOUSE
39	66.5	20.4	377	2	P97327
40	66.5	20.4	377	2	Q920W4
41	66.5	20.4	377	2	Q920W5
42	66.5	20.4	377	2	Q9JL28
43	66.5	20.4	767	1	FTSK_STRPN
44	66.5	20.4	767	1	FTSK_STRR6
45	64	19.6	832	2	Q7YTA5

ALIGNMENTS

RESULT 1	ID	OPUL CRIGR	STANDARD;	PRT;	61 AA.
AC	P83337	2004 (Rel. 45, Last sequence update)			
DT	25-OCT-2004	(Rel. 45, Last annotation update)			
DT	25-OCT-2004	(Rel. 45, Last annotation update)			
DE	GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221) (Peptide O-fucosyltransferase) (O-FucT-1) (Fragment).				
GN	Names=POFUT1;				
OS	Cricetulus griseus (Chinese hamster).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;				
OC	Cricetulus.				
OX	NCBI_TaxID=10029;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;				
RA	Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,				
RA	Haltiwanger R.S.;				
RT	"Modification of epidermal growth factor-like repeats with O-fucose: molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase.";				
RT	J. Biol. Chem. 276:40338-40345(2001).				
RN	[2]				
RP	FUNCTION				
RP	MEDLINE=93364082; PubMed=8358148;				
RA	Harris R.J., Spellman M.W.;				
RT	"O-linked fucose and other post-translational modifications unique to EGF modules.";				
RL	Glycobiology 3:219-224(1993).				
RN	[3]				
RP	FUNCTION.				
RP	MEDLINE=97175972; PubMed=9023546;				
RA	Wang Y., Lee G.P., Kelley R.F., Spellman M.W.;				
RT	"Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains.";				
RL	Glycobiology 6:837-842(1996).				
RN	[4]				
RP	FUNCTION, AND GLYCOSYLATION.				
RP	TISSUE=Ovary;				
RC	MEDLINE=98192601; PubMed=9525914; DOI=10.1074/jbc.273.14.8112;				
RA	Wang Y., Spellman M.W.;				
RT	"Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells.";				
RL	J. Biol. Chem. 273:8112-8118(1998).				
CC	-!- FUNCTION: Catalyzes the reaction that attaches fucose through an O-glycosidic linkage to a conserved serine or threonine residue in EGF domains.				
CC	-!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-beta-L-fucose to the serine hydroxy group of a protein acceptor.				
CC	-!- COFACTOR: Manganese. Other divalent cations increase activity: calcium, cobalt, cadmium, magnesium and nickel.				
CC	-!- PATHWAY: Glycosylation.				

CC -|- PTM: N-glycosylated.
 CC -|- SIMILARITY: Belongs to the glycosyltransferase 68 family.
 KW Direct protein sequencing; Fucose metabolism; Glycoprotein;
 KW Glycosyltransferase; Manganese; Transferase.
 FT NON TER 61
 SQ SEQUENCE 61 AA; 6951 MW; E9507AB60018P23A CRC64;
 Query Match 98.2%; Score 320; DB 1; Length 61;
 Best Local Similarity 100.0%; Pred. No. 3.2e-35; Indels 0; Gaps 0;
 Matches 61; Conservative 0; Mismatches 0;
 QY 1 RLAGSDLAGLYLLYXPXMGREGNOADHFLGSLAFAPKLXVRTILAVPPWIEYQHHPPTNL 60
 DB 1 RLAGSDLAGLYLLYXPXMGREGNOADHFLGSLAFAPKLXVRTILAVPPWIEYQHHPPTNL 60
 QY 61 H 61
 DB 61 H 61
 RESULT 2
 OFUL MOUSE
 ID OFUL MOUSE STANDARD; PRT; 393 AA.
 AC Q91ZW2; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE GDP-fucose protein O-fucosyltransferase 1 precursor (BC 2.4.1.221)
 DE (Peptide O-fucosyltransferase) (O-FUCT-1).
 GN Name=Pofut1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N-3; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -|- FUNCTION: Catalyzes the reaction that attaches fucose through an
 CC O-glycosidic linkage to a conserved serine or threonine residue in
 CC EGF domains (By similarity).
 CC -|- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
 CC beta-L-fucose to the serine hydroxy group of a protein acceptor.

CC -|- COFACTOR: Manganese (By similarity).
 CC -|- PATHWAY: Glycosylation.
 CC -|- SIMILARITY: Belongs to the glycosyltransferase 68 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb.ch/announce/
 CC or send an email to license@isb-sib.ch).
 EMBL; AF375885; AAL09577.1; -;
 EMBL; BC046295; AAL46295.1; -;
 DR MGD; MGI:2153207; Pofut1.
 DR GO; GO:0030173; C:integral to Golgi membrane; ISS.
 DR GO; GO:0008417; F:fucosyltransferase activity; ISS.
 DR GO; GO:0001525; P:angiogenesis; IMP.
 DR GO; GO:0007507; P:heart development; IMP.
 DR GO; GO:0007399; P:neurogenesis; IMP.
 DR GO; GO:0007219; P:Notch signaling pathway; IMP.
 DR GO; GO:0006493; P:O-linked glycosylation; ISS.
 DR GO; GO:0001756; P:somitogenesis; IMP.
 KW Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
 KW Signal; Transferase.
 FT SIGNAL 1 30 Potential.
 FT CHAIN 31 393 GDP-fucose protein O-fucosyltransferase
 FT CARBOHYD 67 67 N-linked (GLCNAC..) (Potential).
 FT CARBOHYD 165 165 N-linked (GLCNAC..) (Potential).
 SQ SEQUENCE 393 AA; 44688 MW; D982104E95E5CF3B CRC64;
 Query Match 93.6%; Score 305; DB 1; Length 393;
 Best Local Similarity 91.8%; Pred. No. 2.3e-32;
 Matches 56; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 RLAGSDLAGLYLLYXPXMGREGNOADHFLGSLAFAPKLXVRTILAVPPWIEYQHHPPTNL 60
 DB 29 RSAGSDLAGLYLLYPCPCMGREGNOADHFLGSLAFAPKLXVRTILAVPPWIEYQHHPPTNL 88
 QY 61 H 61
 DB 89 H 89
 RESULT 3
 O8C8R4
 ID O8C8R4 PRELIMINARY; PRT; 393 AA.
 AC O8C8R4;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,
 DE clone:A930028F21 product:protein O-fucosyltransferase 1, full insert
 DE sequence.
 OS Name=Pofut1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:119-44 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RA "Functional annotation of a full-length mouse cDNA collection.";

Q6E7V0	PRELIMINARY;	PRT;	395 AA.
ID	Q6E7V0		
AC	Q6E7V0;		
DT	25-OCT-2004 (TrEMBLrel. 28, Created)		
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)		
DE	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)		
DE	Protein-O-fucosyltransferase 1.		
GN	Name=ful2;		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OC	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Wistar;		
RX	PubMed=12966037; DOI=10.1093/glycob/cwg113;		
RA	Martinez-Duncker I., Mollicone R., Candellier J.J., Breton C.,		
RA	Oriol R.;		
RT	"A new superfamily of protein-O-fucosyltransferases, alpha2-		
RT	fucosyltransferases and alpha6-fucosyltransferases: phylogeny and		
RT	identification of conserved peptide motifs.";		
RL	Glycobiology 13:1C-5C(2003).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Wistar;		
RA	Martinez-Duncker I., Oriol R., Mollicone R.;		
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AJ781499; CAH03711.1; -.		
DR	GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.		
KW	Glycosyltransferase; Transferase.		
SQ	SEQUENCE 395 AA; 44983 MW; 2A85521E22F38920 CRC64;		
	Query Match 92.6%; Score 302; DB 2; Length 395;		
	Best Local Similarity 90.2%; Pred. No. 5.9e-32;		
	Matches 55; Conservative 0; Mismatches 6; Indels 0; Gaps 0;		
Qy	1 RLAGSWDLAGLYLXXPMGRFGNQADHFLGSLAFAKLVKRTLAVPPMIEYOHKPPFTNL 60		
Db	31 RAGSWDLAGLYLYCPCMRGRFGNQADHFLGSLAFAKLVKRTLAVPPMIEYOHKPPFTNL 90		
Qy	61 H 61		
Db	91 H 91		
RESULT 5			
OFUL_HUMAN	STANDARD;	PRT;	388 AA.
ID	OFUL_HUMAN		
AC	Q9H488; Q14685; Q9BW76;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DE	25-OCT-2004 (Rel. 45, Last annotation update)		
DE	GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)		
DE	(Peptide O-fucosyltransferase) (O-FucT-1).		
GN	Name=FOFUT1; Synonyms=KIAA0180;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OC	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.		
RC	TISSUE=Heart;		
RX	MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;		
RA	Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,		
RA	Haltiwanger R.S.;		
RT	"Modification of epidermal growth factor-like repeats with O-fucose:		
RT	molecular cloning and expression of a novel GDP-fucose protein O-		
RT	fucosyltransferase.";		
RL	J. Biol. Chem. 276:40338-40345(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	TISSUE=Brain;		
RX	MEDLINE=22158633; PubMed=12168954;		

DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
GN (Peptide O-fucosyltransferase) (O-FucT-1).
OS Name=POFUT1; Synonyms=FUT12;
OC Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RI Oriol R.;
RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT identification of conserved peptide motifs.";
RL Glycobiology 13:1C-5C(2003).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains (By similarity).
CC -!- CATALYTIC ACTIVITY: transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese (By similarity).
CC -!- PATHWAY: Glycosylation.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DE EMBL: AJ781500; CAH03712.1; -;
KW Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW Signal; Transferase. 26
FT SIGNAL 1 Potential.
FT CHAIN 27 388 GDP-fucose protein O-fucosyltransferase
FT 1.
FT CARBOHYD 62 62 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 388 AA; 43955 MW; 3FACCA434D02415 CRC64;

Query Match 90.5%; Score 295; DB 1; Length 388;
Best Local Similarity 91.5%; Pred. No. 5.1e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 26 AGSWDPAGYLLYCPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 84

RESULT 7
QYRE6 PRELIMINARY; PRT; 351 AA.
AC QYRE6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Protein O-fucosyltransferase 1b.
GN Name=POFUT1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX Liorol C., Germot A., Dupuy F., Maftah A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: AY344581; AAQ02333.1; -;
DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.

KW Glycosyltransferase; Transferase.
SQ SEQUENCE 351 AA; 39576 MW; BF830F61A7296F42 CRC64;

Query Match 88.3%; Score 288; DB 2; Length 351;
Best Local Similarity 89.7%; Pred. No. 4e-30; 5; Indels 0; Gaps 0;
Matches 52; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GSWDLAGYLLYXPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 30 GSWDPAGYLLYCPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNVH 87

RESULT 8

QYRE7 PRELIMINARY; PRT; 391 AA.
AC QYRE7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Protein O-fucosyltransferase 1a.
GN Name=POFUT1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX Liorol C., Germot A., Dupuy F., Maftah A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY344580; AAQ02332.1; -;
DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 391 AA; 44307 MW; 96762AB81A2027AD CRC64;

Query Match 88.3%; Score 288; DB 2; Length 391;
Best Local Similarity 89.7%; Pred. No. 4.5e-30;
Matches 52; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 GSWDLAGYLLYXPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 30 GSWDPAGYLLYCPXMGFRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNVH 87

RESULT 9

QYRE3 PRELIMINARY; PRT; 378 AA.
AC QYRE3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Protein-O-fucosyltransferase (Fragment).
GN Name=POFUT1;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
RT identification of conserved peptide motifs.";
RL Glycobiology 13:1C-5C(2003).
DR EMBL: AJ567917; CAD99201.1; -;
DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.

KW Glycosyltransferase; Transferase.
FT NON TER 1
FT NON TER 378
SQ SEQUENCE 378 AA; 42842 MW; 1870DEB0C50713F1 CRC64;

Query Match 87.7%; Score 286; DB 2; Length 378;
Best Local Similarity 89.8%; Pred. No. 8e-30;


```

OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Duncker I., Oriol R., Mollicone R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ781498; CAH03710.1; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 396 AA; 45056 MW; F90504E19507073F CRC64;

Query Match 76.4%; Score 250; DB 2; Length 396;
Best Local Similarity 75.0%; Pred.No. 5.8e-25;
Matches 42; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 6 WDLAGLYLXXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 WDSGGYLYPCMGRCFGNQADHFLGSLAFKLVRTLAVPPWIVYVYVYVYVYVYV 92

RESULT 14
Q7T028 PRELIMINARY; PRT; 395 AA.
AC Q7T028;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221).
GN Namepofut1; Synonymys=fut12;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Duncker I., Mollicone R., Candellier J.J., Breton C.,
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ579536; CAE18459.1; -.
DR ZFIN; ZDB-GENE-040303-2; Pofut1.
DR GO; GO:0046922; F:peptide-O-fucosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 395 AA; 45165 MW; 15BBE06172542E8A CRC64;

Query Match 76.4%; Score 249; DB 2; Length 395;
Best Local Similarity 75.4%; Pred.No. 7.9e-25;
Matches 43; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 5 SWDLAGLYLXXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 33 TWDENGHLYLPCMGRCFGNQADHFLGSLAFKLVRTLAVPPWIVYVYVYVYVYVYV 89

RESULT 15
Q70AG7 PRELIMINARY; PRT; 384 AA.
AC Q70AG7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Protein-O-fucosyltransferase.
GN Name=fut12;
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Martinez-Duncker I., Mollicone R., Candellier J.J., Oriol R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ606070; CAE54305.1; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR InterPro; IPR000886; ER_target_S.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 384 AA; 43650 MW; 7CAA6ADC90F40032 CRC64;

Query Match 75.8%; Score 247; DB 2; Length 384;
Best Local Similarity 76.8%; Pred.No. 1.4e-24;
Matches 43; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 6 WDLAGLYLXXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 23 WDQNGVLYLPCMGRCFGNQADHFLGSLAFKLVRTLAVPPWIVYVYVYVYVYVYV 78

Search completed: October 25, 2005, 15:35:24
Job time : 36.1971 secs

```

THIS PAGE BLANK (USPTO)

Against SNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:59:38 ; Search time 1065.43 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGLLYXPXGR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2 1/USPTO.spool h/US09774954/tunat 25102005 105432 6406/app query.fasta_1.917
-DB=EST -QPM=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.1 5334 @runat 25102005 105432 6406 -NCPU=6 -ICPU=3
-MM MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gse1:*

9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	305	93.6	318	6	BY778150
2	305	93.6	333	5	BY175344
3	305	93.6	352	5	BY328132
4	305	93.6	456	2	BB633858
5	305	93.6	480	2	BB851550
6	305	93.6	536	7	CF155036
7	305	93.6	604	2	BB618795
8	305	93.6	605	6	BY728018
9	305	93.6	611	2	BB623905

10	305	93.6	614	6	BY752500
11	305	93.6	634	2	BB631131
12	305	93.6	641	2	BB611933
13	305	93.6	650	6	BY752497
14	305	93.6	661	2	BB642654
15	305	93.6	664	6	BY728337
16	305	93.6	695	2	BB617910
17	305	93.6	915	5	BQ936634
18	305	93.6	947	4	BI410783
19	305	93.6	988	5	BQ964179
20	305	93.6	2413	3	AK037857
21	305	93.6	2616	3	AK044629
22	302	92.6	763	7	CK366990
23	302	92.6	3611	3	AK081059
24	298	91.4	304	5	BY322240
25	295	90.5	520	7	CR543310
26	295	90.5	550	7	CR348927
27	295	90.5	551	2	BE261136
28	295	90.5	576	2	BE262956
29	295	90.5	595	7	CK348926
30	295	90.5	610	7	AL555719
31	295	90.5	656	7	CV029252
32	295	90.5	699	1	AU137536
33	295	90.5	705	7	CK348925
34	295	90.5	753	4	BG424731
35	295	90.5	790	4	BG519639
36	295	90.5	808	4	BG747760
37	295	90.5	841	5	BK345989
38	295	90.5	889	4	B1524031
39	295	90.5	898	2	BF315757
40	295	90.5	919	5	BUI55142
41	295	90.5	942	7	CO774833
42	295	90.5	950	4	BI199190
43	295	90.5	955	2	BE260030
44	295	90.5	1555	3	CR617950
45	295	90.5	1711	2	BF206548

ALIGNMENTS

RESULT 1

BY778150

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY778150 RIKEN full-length enriched, 17.5 days embryo whole body
Mus musculus cDNA clone L930123D17 5', mRNA sequence.
318 bp mRNA linear EST 23-MAR-2004

BY778150.1 GI:39704789

Mus musculus (house mouse)

EST.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 318)

Carninci, P., Waki, K., Shiraki, T., Konno, H., Shibata, K., Itoh, M.,

Aizawa, K., Arakawa, T., Ishii, Y., Sasaki, D., Bono, H., Kondo, S.,

Sugahara, Y., Saito, R., Osato, N., Fukuda, S., Sato, K., Watahiki, A.,

Hirozane-Kishikawa, T., Nakamura, M., Shibata, Y., Yasunishi, A.,

Kikuchi, N., Yoshiki, A., Kusakabe, M., Gustincich, S., Beisel, K.,

Pavan, W., Aidinis, V., Nakagawara, A., Held, W. A., Iwata, H., Kono, T.,

Nakauchi, H., Lyons, P., Wells, C., Hume, D. A., Fagiolini, M.,

Hensch, T. K., Brinkmeier, M., Camper, S., Hirota, J., Mombaerts, P.,

Muranatsu, M., Okazaki, Y., Kawai, J. and Hayashizaki, Y.

Targeting a complex transposon: the construction of the mouse

full-length cDNA encyclopedia

Genome Res. 13 (6B), 1273-1289 (2003)

22703353

12819125

CONTACT: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.jp/) for
 further details.

FEATURES

Location/Qualifiers
 1..338
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="L930123P17"
 /tissue_type="whole body"
 /dev_stage="17.5 days embryo"
 /clone_lib="RIKEN full-length enriched, 17.5 days embryo
 whole body"

ORIGIN

Alignment Scores: 318
 Pred. No.: 9,556-32 Length: 56
 Score: 105.00 Matches: 56
 Percent Similarity: 91.80% Conservative: 0
 Best Local Similarity: 91.80% Mismatches: 5
 Query Match: 93.56% Indels: 0
 DB: 6 Gaps: 0

US-09-774-954-3 (1-61) x BY778150 (1-318)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTy*****MetGlyArg 20
 DB 115 CGTCAAGCGGCTCTGGAGCTCGCGGTACTCTGCTCTCTGCTCGATCGGGCGC 174
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
 DB 175 TTTGGGAACCGAGCTGATCATCTTCTGGCTCCGTGGCATTTGGCAAGCTGTGAACCCG 234
 QY 41 ThrLeuAlaValProProTrrpIleGlyTyrGlnHisLysProPropHeThrAsnLeu 60
 DB 235 ACCTTGGCTGTACCTCCATGGATTGATACCAATCAATCAAGCCTCTTTCAACCACTC 294
 QY 61 His 61
 DB 295 CAT 297

RESULT 2
 BY175344
 LOCUS BY175344 RIKEN full-length enriched, NOD-derived CD11c +ve
 DEFINITION dendritic cells Mus musculus cDNA Clone F630011F24 5', mRNA
 sequence.
 BY175344
 BY175344.1 GI:26311990
 EST.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 333)

Okazaki.Y., Furuno.M., Kasukawa.T., Adachi.J., Bono.H., Kondo.S.,

Nikaido.I., Osato.N., Saito.R., Suzuki.H., Yamanaka.I.,

Kiyosawa.H., Yagi.K., Tomaru.Y., Hasegawa.Y., Nogami.A.,

Schonbach.C., Gojobori.T., Baldarelli.R., Hill.D.P., Bult.C.,

Hume.D.A., Quackenbush.J., Schriml.L.M., Kanapin.A., Matsuda.H.,

Batalov.S., Beisel.K.W., Blake.J.A., Bradt.D., Brusci.V.,

Ciochia.C., Corbani.L.E., Cousins.S., Dalla.E., Dragani.T.A.,

Fletcher.C.P., Forrest.A., Frazer.K.S., Gaasterland.T.,

Gariboldi.M., Gissi.C., Godzik.A., Gough.J., Grimmond.S.,

Gustinch.S., Hirokawa.N., Jackson.I.J., Jarvis.E.D., Kanai.A.,

Kawaji.H., Kawasaki.Y., Kedzierski.R.M., King.B.L., Konagaya.A.,

Kurochkin.I.V., Lee.Y., Lenhard.B., Lyons.P.A., Maglott.D.R.,
 Maltais.L., Marchionni.L., McKenzie.L., Miki.H., Nagashima.T.,
 Numata.K., Okido.T., Pavan.W.J., Pertea.G., Pesole.G.,
 Petrovsky.N., Pillai.R., Pontius.J.U., Qi.D., Ramachandran.S.,
 Ravasi.T., Reed.J.C., Reed.D.J., Reid.J., Ring.B.Z., Ringwald.M.,
 Sandelin.A., Schneider.C., Semple.C.A., Setou.M., Shimada.K.,
 Sultana.R., Takekura.Y., Taylor.M.S., Teasdale.R.D., Tomita.M.,
 Verardo.R., Wagner.L., Wahlestedt.C., Wang.Y., Watanabe.Y.,
 Wells.C., Wilming.L.G., Wynshaw-Boris.A., Yanagisawa.M., Yang.I.,
 Yang.L., Yuan.Z., Zavolan.M., Zhu.Y., Zimmer.A., Carninci.P.,
 Hayatsu.N., Hirozane-Kishikawa.T., Konno.H., Nakamura.M.,
 Sakazume.N., Sato.K., Shiraki.T., Waki.K., Kawai.J., Aizawa.K.,
 Arakawa.T., Fukuda.S., Hara.A., Hashizume.W., Imotani.K., Ishii.Y.,
 Itoh.M., Kagawa.I., Miyazaki.A., Sakai.K., Sasaki.D., Shibata.K.,
 Shingawa.A., Yasunishi.A., Yoshino.M., Waterston.R., Lander.E.S.,
 Rogers.J., Birney.E. and Hayashizaki.Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa.K., Akimura.T., Arakawa.T., Carninci.P., Fukuda.S.,
 Hirozane.T., Imotani.K., Ishii.Y., Itoh.M., Kawai.J., Konno.H.,
 Miyazaki.A., Murata.M., Nakamura.M., Nomura.K., Numazaki.R.,
 Ohno.M., Sakai.K., Sakazume.N., Sasaki.D., Sato.K., Shibata.K.,
 Shiraki.T., Tagami.M., Waki.K., Watahiki.A., Muramatsu.M. and
 Hayashizaki.Y. Direct Submission
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630. (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

Location/Qualifiers

1..333
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="taxon:10090"
 /clone="F630011F24"
 /cell_type="NOD-derived CD11c +ve dendritic cells"
 /clone_lib="RIKEN full-length enriched, NOD-derived CD11c
 +ve dendritic cells"

ORIGIN

Alignment Scores:

Pred. No.: 1,016-31 Length: 333

Score: 305.00 Matches: 57

Percent Similarity: 93.44% Conservative: 0

Best Local Similarity: 93.44% Mismatches: 4
 Query Match: 93.56% Indels: 0
 DB: 5 Gaps: 0

US-09-774-954-3 (1-61) x BY175344 (1-333)

Qy 1 AtgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
 Db 91 CGGTACGGGGCTCTGGAGACCTGGCGGTACTCTACTGCTCTCTGTCATGGGGCGC 150
 Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
 Db 151 TTTGGGAACAGGCTGATCATTCTTTGGGCTCCCTGGCATTTTCGAAGCTGTGTAACCGC 210
 Qy 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProPheThrAsnLeu 60
 Db 211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCTTTCACCAACCTC 270
 Qy 61 His 61
 Db 271 CAT 273

RESULT 3

BY328132 352 bp mRNA linear EST 11-DEC-2002
 BY328132 RIKEN full-length enriched, synovial fibroblasts Mus
 LOCUS
 DEFINITION
 musculus cDNA clone LI30001N19 5', mRNA sequence.

ACCESSION

VERSION
 BY328132
 EST.

KEYWORDS

SOURCE
 Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Kiyosawa, H., Yagi, K., Osato, N., Saito, R., Suzuki, H., Yamana, I.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Sult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godrik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verdardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wyehaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 JOURNAL
 MEDLINE
 PUBMED
 22354683
 12468851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel.: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Vassilis Aidinis (Biomedical Sciences
 Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
 Fleming street 16672 Vari, Greece) whose assistance we gratefully
 acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

1. .352
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="LI30001N19"
 /cell_type="synovial fibroblasts"
 /clone_lib="RIKEN full-length enriched, synovial
 fibroblasts"

ORIGIN

Alignment Scores:

Pred. No.: 1,08e-31 Length: 352
 Score: 305.00 Matches: 56
 Percent Similarity: 91.80% Conservative: 0
 Best Local Similarity: 91.80% Mismatches: 5
 Query Match: 93.56% Indels: 0
 DB: 5 Gaps: 0

US-09-774-954-3 (1-61) x BY328132 (1-352)

Qy 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
 Db 112 CGGTACGGGGCTCTGGAGACCTGGCGGTACTCTACTGCTCTGTCATGGGGCGC 171
 Qy 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
 Db 172 TTTGGGAACAGGCTGATCATTCTTTGGGCTCCCTGGCATTTGCGAAGCTGTGAACCGC 231
 Qy 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProPheThrAsnLeu 60
 Db 232 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCTTTCACCAACCTC 291
 Qy 61 His 61
 Db 292 CAT 294

RESULT 4

BB633858
 LOCUS
 DEFINITION
 musculus cDNA clone A330084M14 5', mRNA sequence.
 BB633858 RIKEN full-length enriched, adult male spinal cord Mus
 BB633858
 ACCESSION

Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

source

Location/Qualifiers

1. .480
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370002F02"
 /cell_type="B16 F10Y cells"
 /clone_lib="RIKEN full-length enriched, B16 F10Y cells"

ORIGIN

Alignment Scores:
 Pred. No.: 1.58e-31 Length: 480
 Score: 305.00 Matches: 56
 Percent Similarity: 91.80% Conservative: 0
 Best Local Similarity: 91.80% Mismatches: 5
 Query Match: 93.56% Indels: 0
 DB: 2 Gaps: 0

US-09-774-954-3 (1-61) x BB851550 (1-480)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
 |||||
 Db 117 CGGTACGGGGCTCTCGGACCTCGCGGTACTGCTACTGCTCCATGCGGGCGC 176
 |||||
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
 |||||
 Db 177 TTGTGGGAACAGCGCTGATCATCTTCTTGGGCTCCGTCATTTGCGAAGCTGTGAACCGC 236
 |||||
 QY 41 ThrLeuAlaValProProTrpPilleGluTyrGlnHisLysProProPheThrAsnLeu 60
 |||||
 Db 237 ACCTTGCTGTACTCCATGGATTGAATACCAACATCACAGGCTCTCTTCCACCACTC 296

QY 61 His 61

Db 297 CAT 299

RESULT 6

CF155036

LOCUS

CF155036 536 bp mRNA linear EST 25-JUL-2003
 B0620C09-5 NIA Mouse Blastocyst cDNA Library (Long) Mus musculus

cDNA clone NIA:B0620C09 IMAGE:30450176 5', mRNA sequence.

ACCESSION

CF155036

VERSION

CF155036.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

AUTHORS

Piao,Y., KO,N.T., Lim,M.K. and Ko,M.S.H.

TITLE

Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification

method

Genome Res. 11 (9), 1553-1558 (2001)

JOURNAL

MEDLINE

21429098

PUBMED

11544199

COMMENT

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@lgsun.grc.nia.nih.gov

Plate: B0620 row: C column: 09

Seq primer: M13 Reverse

High quality sequence stop: 536

POLYA=No.

FEATURES

source

Location/Qualifiers

1. .536
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="niaEST:B0620C09-5"
 /db_xref="taxon:10090"
 /clone="NIA:B0620C09 IMAGE:30450176"
 /tissue_type="Blastocyst"
 /dev_stage="3.5-dpc"
 /lab_host="PH10B"
 /clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
 /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun-grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen].
 5'-pGACATGTTCTAGATCGGCGCGCCCTTTTCTTTT-3' from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN

Alignment Scores:
 Pred. No.: 1.8e-31 Length: 536
 Score: 305.00 Matches: 56
 Percent Similarity: 91.80% Conservative: 0
 Best Local Similarity: 91.80% Mismatches: 5
 Query Match: 93.56% Indels: 0
 DB: 7 Gaps: 0

US-09-774-954-3 (1-61) x CF155036 (1-536)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
 |||||
 Db 84 CGGTACGGGGCTCTCGGACCTCGCGGTACTGCTACTGCTCCATGCGGGCGC 143
 |||||
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
 |||||
 Db 144 TTGTGGGAACAGCGCTGATCATCTTCTTGGGCTCCGTCATTTGCGAAGCTGTGAACCGC 203
 |||||
 QY 41 ThrLeuAlaValProProTrpPilleGluTyrGlnHisLysProProPheThrAsnLeu 60
 |||||
 Db 204 ACCTTGCTGTACTCCATGGATTGAATACCAACATCACAGGCTCTTTCACCACTC 263

QY 61 His 61

Db 264 CAT 266

RESULT 7

BB618795

LOCUS

DEFINITION

BB618795

604 bp mRNA linear

EST 31-AUG-2001

RIKEN full-length enriched, 8 days embryo Mus musculus

CDNA clone 5730432G02 5', mRNA sequence.

BB618795
BB618795.1 GI:15396691
EST.
Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 604)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Onno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
Location/Qualifiers
1. .604
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730432G02"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGTCCACAGAGCTCTTTTATTTTNN 3'], cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to

Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

ORIGIN
Alignment Scores:
Pred. No.: 2,09e-31 Length: 604
Score: 305.00 Matches: 56
Percent Similarity: 91.80% Conservative: 0
Best Local Similarity: 91.80% Mismatches: 5
Query Match: 93.56% Indels: 0
DB: 2 Gaps: 0
US-09-774-954-3 (1-61) x BB618795 (1-604)
QY 1 ArgLeuAlaGlySerTTPAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 110 CGGTCAGCGGGCTCTCGGACCTGGCGGTTACTCTCTACTCTCGCTGGCGGC 169
QY 21 PheGlyValAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeuLeu***ValArg 40
Db 170 TTTGGGAACCGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTCTGAACCGC 229
QY 41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisHisLeuProProPheThrAsnLeu 60
Db 230 ACCTTGGCTGTACTCCATCGATGATGAATCAACATCAACAGCTCTCTTCCACCACTC 289
QY 61 His 61
Db 290 CAT 292
RESULT 8
BY728018 605 bp mRNA linear EST 17-DEC-2002
LOCUS BY728018 RIKEN full-length enriched, 10 days neonate cerebellum Mus
DEFINITION musculus cDNA clone B930076G19 5', mRNA sequence.
ACCESSION BY728018
VERSION BY728018.1 GI:27141145
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 605)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaado, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusci, V., Ciothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlested, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hiroxane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Haseizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,

Score: 305.00 Matches: 56
Percent Similarity: 91.80% Conservative: 0
Best Local Similarity: 91.80% Mismatches: 5
Query Match: 93.56% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-3 (1-61) x BB611131 (1-634)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 104 CGGTACGGGGCTCTGGGACCTGGCGGTACCTGCTCTACTGCTCCCTGCATGGGGCGC 163
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
Db 164 TTGGGAACAGGCTGATCACTCTTGGGCTCCCTGGCATTTGGCAAGCTGCTGAACCGC 223
QY 41 ThrLeuAlaValProProTyrPheGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 224 ACCTTGGCTGTACCTCCATGATTAATCAACATCAACAGCTCTCTTCCACCACTC 283
QY 61 His 61
Db 284 CAT 286

RESULT 12

BB611933 641 bp mRNA linear EST 26-OCT-2001
BB611933 RIKEN full-length enriched, 15 days embryo head Mus
musculus cDNA clone 4021401D01 5', mRNA sequence.

BB611933.1 GI:16453020

Mus musculus (house mouse)

EST.

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES

Location/Qualifiers
1..641
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4021401D01"
/sex="mixed"
/tissue_type="head"
/dev_stage="15 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 15 days embryo
head"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCACAGAGCTTTTCTTTTCTTNN 3']. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAATAATATCCCTCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I"

ORIGIN

Alignment Scores:
Pred. No.: 2,249-31 Length: 641
Score: 305.00 Matches: 56
Percent Similarity: 91.80% Conservative: 0
Best Local Similarity: 91.80% Mismatches: 5
Query Match: 93.56% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-3 (1-61) x BB611933 (1-641)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 91 CGGTACGGGGCTCTGGGACCTGGCGGTACCTGCTCTACTGCTCCCTGCATGGGGCGC 150
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
Db 151 TTGGGAACAGGCTGATCACTCTTGGGCTCCCTGGCATTTGGCAAGCTGCTGAACCGC 210
QY 41 ThrLeuAlaValProProTyrPheGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 211 ACCTTGGCTGTACCTCCATGATTAATCAACATCAACAGCTCTCTTCCACCACTC 270
QY 61 His 61
Db 271 CAT 273

RESULT 13

BY752497 650 bp mRNA linear EST 17-DEC-2002
LOCUS BY752497 RIKEN full-length enriched, 9.5 days embryo parthenogenote
DEFINITION Mus musculus cDNA clone B13000118 5', mRNA sequence.
ACCESSION BY752497
VERSION BY752497.1 GI:27183400
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y. Riken Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.jp, URL: http://genome.gsc.riken.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Koneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y. Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct. Funct. Genomics* 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers

1. 661
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A930028P21"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult retina"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATAATTAATCCCTCCCTCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pluscript KS(+) after bulk excision from Lambda FIC I. Retina RNA was provided by Stefano Guatini, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA, whose assistance we gratefully acknowledge."

ORIGIN

Alignment Scores: 2.33e-31 Length: 661
Pred. No.: 305.00 Matches: 56
Score: 91.80% Conservative: 0

Best Local Similarity: 91.80% Mismatches: 5
Query Match: 93.56% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-3 (1-61) x BB642654 (1-661)

QY 1 ArgLeuAlaGlySerTTPspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
118 CGGTGACGGGGTCTCTGGGACCTGGCGGTACTCTCTACTGTCCCTGCAATGGGCGC 177
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
178 TTGGGAACGAGGCTCATCACTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 237
QY 41 ThrLeuAlaValProProTyrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
238 ACCTTGGCTGTACTCCATGATGGAATACCAACATCACAGGCTCTCTTTCCACCACTC 297
QY 61 His 61
DB 298 CAT 300

RESULT 15
BY728337
LOCUS
DEFINITION
BY728337 RIKEN full-length enriched, 16 days embryo head Mus
musculus cDNA clone Cl30003G23 5', mRNA sequence.
ACCESSION
BY728337
VERSION
BY728337.1 GI:27141464
KEYWORDS
EST.
SOURCE
Mus musculus
ORGANISM
Mus musculus (house mouse)
REFERENCE
1 (bases 1 to 664)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Wagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusci, V., Ciothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Guatini, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lennard, B., Lyons, P. A., Maglott, D. R., Maltas, D., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
NATURE 420, 563-573 (2002)
JOURNAL
MEDLINE
PUBMED
2354683
12466851
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 : Search time 260.659 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGLYLXPXWGR.....LAVPPWIEYQHKKPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9772377 seqs, 4126317084 residues

Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09774954/runat_25102005.105434.6433/app_query.fasta_1.917
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09774954 @cgn 1 1 684 @runat_25102005.105434_6433
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=1 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:
18: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq:
19: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:
20: /cgn2_6/prodata/1/pubpna/US10G_PUBCOMB.seq:
21: /cgn2_6/prodata/1/pubpna/US10H_PUBCOMB.seq:
22: /cgn2_6/prodata/1/pubpna/US10I_PUBCOMB.seq:
23: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
24: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
25: /cgn2_6/prodata/1/pubpna/US11A_PUBCOMB.seq:
26: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:
27: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
28: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	295	90.5	1100	11	US-09-774-954-16	Sequence 16, Appl
2	295	90.5	1300	11	US-09-774-954-4	Sequence 4, Appl
3	295	90.5	1400	22	US-10-956-157-9986	Sequence 9986, Ap
4	295	90.5	1506	22	US-10-956-157-4751	Sequence 4751, Ap
5	295	90.5	1514	11	US-09-774-954-1	Sequence 1, Appl
6	295	90.5	5218	16	US-10-301-822-168	Sequence 168, App
7	295	90.5	5218	22	US-10-956-157-4752	Sequence 4752, Ap
8	295	90.5	11284	11	US-09-774-954-5	Sequence 5, Appl
9	204	62.6	5009	11	US-09-774-954-7	Sequence 7, Appl
10	177	54.3	1209	26	US-11-097-143-18161	Sequence 18161, A
11	145	44.5	3264	26	US-11-097-143-18160	Sequence 18160, A
12	145	44.5	3793	26	US-11-097-143-18082	Sequence 18082, A
13	76	23.3	521	13	US-09-925-065A-602444	Sequence 602444,
14	76	23.3	760	24	US-10-820-474A-236	Sequence 236, App
15	72.5	22.2	1803	16	US-10-156-761-5001	Sequence 5001, Ap
16	72.5	22.2	9025608	16	US-10-156-761-1	Sequence 1, Appl
17	71.5	21.9	568	21	US-10-425-115-16666	Sequence 16666, A
18	69.5	21.3	594	18	US-10-282-122A-25606	Sequence 25606, A
19	69	21.2	764	21	US-10-425-115-28075	Sequence 28075, A
20	68.5	21.0	829	19	US-10-424-599-123276	Sequence 123276,
21	68	20.9	4953	20	US-10-437-963-23967	Sequence 23967, A
22	67.5	20.7	1617	20	US-10-437-963-11436	Sequence 11436, A
23	66.5	20.4	636	13	US-09-925-065A-285729	Sequence 285729,
24	66.5	20.4	2232	22	US-10-472-928-1697	Sequence 1697, Ap
25	66.5	20.4	2232	24	US-10-617-320-1007	Sequence 1007, Ap
26	66.5	20.4	2304	9	US-09-815-242-9245	Sequence 9245, Ap
27	66.5	20.4	2304	9	US-09-815-242-9500	Sequence 9500, Ap
28	66.5	20.4	2304	10	US-09-769-787-308	Sequence 308, App
29	66.5	20.4	2304	18	US-10-282-132A-37727	Sequence 37727, A
30	66.5	20.4	2352	9	US-09-775-978-1	Sequence 1, Appl
31	66.5	20.4	18627	8	US-08-961-527-113	Sequence 113, App
32	66.5	20.4	18627	18	US-10-158-844-113	Sequence 113, App
33	66.5	20.2	2162598	22	US-10-472-928-4979	Sequence 4799, Ap
34	66	20.2	361	13	US-09-925-065A-476650	Sequence 476650,
35	65	19.9	478	19	US-10-152-319A-21	Sequence 21, Appl
36	65	19.9	519	20	US-10-437-963-80416	Sequence 80416, A
37	65	19.9	1437	19	US-10-424-599-40441	Sequence 40441, A
38	65	19.9	2910	19	US-10-424-599-40437	Sequence 40437, A
39	65	19.9	513509	10	US-09-754-853A-4	Sequence 4, Appl
40	64.5	19.8	476	13	US-09-925-065A-500787	Sequence 500787,
41	64.5	19.8	476	13	US-09-925-065A-500788	Sequence 500788,
42	64.5	19.8	575	24	US-10-631-467-1068	Sequence 1068, Ap
43	64.5	19.8	628	13	US-09-925-065A-770388	Sequence 770388,
44	64	19.6	600	24	US-10-972-079-32895	Sequence 32895, A
45	64	19.6	1018	18	US-10-282-122A-36872	Sequence 36872, A

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

2

```

DB:                22                Gaps:                0
US-09-774-954-3 (1-61) x US-10-956-157-9986 (1-1400)
QY  3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db  19 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCGCCCTGCATGGGGCGCTTTGGG 78
QY  23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db  79 AACGAGCGCGATCATTCTTGGGCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 138
QY  43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db  139 GCTGTCCCTCTTGGATTGATTGATACGAGCATCACAGCCCTCTTTACCAACCTCCAT 195

RESULT 4
US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; NUMBER OF SEQ ID NOS: 2004-10-04
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 4751
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:
Pred. No.: 5.24e-33 Length: 1506
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-3 (1-61) x US-10-956-157-4751 (1-1506)
QY  3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db  125 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCGCCCTGCATGGGGCGCTTTGGG 184
QY  23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db  185 AACGAGCGCGATCATTCTTGGGCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 244
QY  43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db  245 GCTGTCCCTCTTGGATTGATTGATACGAGCATCACAGCCCTCTTTACCAACCTCCAT 301

RESULT 5
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-Nov-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-Jan-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Alignment Scores:
Pred. No.: 5.27e-33 Length: 1514
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-3 (1-61) x US-09-774-954-1 (1-1514)
QY  3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db  7 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCGCCCTGCATGGGGCGCTTTGGG 66
QY  23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db  67 AACGAGCGCGATCATTCTTGGGCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 126
QY  43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db  127 GCTGTCCCTCTTGGATTGATTGATACGAGCATCACAGCCCTCTTTACCAACCTCCAT 183

RESULT 6
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10

```

;; PRIOR APPLICATION NUMBER: US 60/361,978
;; PRIOR FILING DATE: 2002-03-05
;; PRIOR APPLICATION NUMBER: US 60/381,988
;; PRIOR FILING DATE: 2002-05-20
;; NUMBER OF SEQ ID NOS: 228
;; SOFTWARE: PastSeq for Windows Version 4.0
;; SEQ ID NO 168
;; LENGTH: 5218
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (50)....(1216)
US-10-301-822-168

Alignment Scores: 2,18e-32 Length: 5218
Pred. No.: 295.00 Matches: 54
Score: 295.00
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 16 Gaps: 0

US-09-774-954-3 (1-61) x US-10-301-822-168 (1-5218)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
DB 125 GCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGCTTTGGG 184
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
DB 185 AACCGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 244
QY 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
DB 245 GCTGTCCCTCTGGATTGAGTACCAAGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 301

RESULT 7

US-10-956-157-4752
;; Sequence 4752, Application US/10956157
;; Publication No. US20050118623A1
;; GENERAL INFORMATION:
;; APPLICANT: Wyeth
;; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;; FILE REFERENCE: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;; CURRENT APPLICATION NUMBER: US/10/956,157
;; CURRENT FILING DATE: 2004-10-04
;; NUMBER OF SEQ ID NOS: 319805
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 4752
;; LENGTH: 5218
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores: 2,18e-32 Length: 5218
Pred. No.: 295.00 Matches: 54
Score: 295.00
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-3 (1-61) x US-10-956-157-4752 (1-5218)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
DB 125 GCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGCTTTGGG 184
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42

DB 185 AACCGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 244
QY 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
DB 245 GCTGTCCCTCTGGATTGAGTACCAAGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 301
RESULT 8
US-09-774-954-5
;; Sequence 5, Application US/09774954
;; Publication No. US20040241645A1
;; GENERAL INFORMATION:
;; APPLICANT: Yang Wang, Michael W. Spellman
;; TITLE OF INVENTION: O-Fucosyltransferase
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 1 DNA Way
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WinPatIn (Genentech)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/774,954
;; FILING DATE: 30-Jan-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/978,741
;; FILING DATE: 26-NOV-1997
;; APPLICATION NUMBER: 08/792,498
;; FILING DATE: 31-JAN-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Svoboda, Craig G.
;; REGISTRATION NUMBER: 39,044
;; REFERENCE/DOCKET NUMBER: P1041P1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/225-1489
;; TELEFAX: 650/952-9881
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 11284 base pairs
;; TYPE: Nucleic Acid
;; STRANDEDNESS: Single
;; TOPOLOGY: Linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

Alignment Scores: 5,29e-32 Length: 11284
Pred. No.: 295.00 Matches: 54
Score: 295.00
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-3 (1-61) x US-09-774-954-5 (1-11284)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
DB 4242 GCGGGCTCTGGGACCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGCTTTGGG 4301
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
DB 4302 AACCGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 4361
QY 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
DB 4362 GCTGTCCCTCTTGGATTGAGTACCAAGCTCTCTTTCACCAACCTCCAT 4418


```
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 18160
/ LENGTH: 3264
/ TYPE: DNA
/ ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.: 1.58e-10 Length: 3264
Score: 145.00 Matches: 26
Percent Similarity: 87.88% Conservative: 3
Best Local Similarity: 78.79% Mismatches: 4
Query Match: 44.48% Indels: 0
DB: 26 Gaps: 0

US-09-774-954-3 (1-61) x US-11-097-143-18160 (1-3264)
QY 19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
DB 2092 GGACGCTTTGGCAACAGCGCGACCACTTCCTGGGATCATTTGGCCTTCGCCAAGGCGCTT 2033
QY 39 ValArgThrLeuAlaValProTfPileGluTyrGln 51
DB 2032 AATCGCACCTGATCTCCGCCCGTGGTGGATATCGT 1994

RESULT 12
/ Sequence 18082, Application US/11097143
/ Publication No. US20050208558A1
/ GENERAL INFORMATION:
/ APPLICANT: Venter, J. Craig
/ TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
/ TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
/ FILE REFERENCE: CL000728
/ CURRENT APPLICATION NUMBER: US/11/097,143
/ PRIOR FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: 60/157,832
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: 60/160,191
/ PRIOR FILING DATE: 1999-10-19
/ PRIOR APPLICATION NUMBER: 60/161,932
/ PRIOR FILING DATE: 1999-10-28
/ PRIOR APPLICATION NUMBER: 60/164,769
/ PRIOR FILING DATE: 1999-11-12
/ PRIOR APPLICATION NUMBER: 60/173,383
/ PRIOR FILING DATE: 1999-12-28
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 18082
/ LENGTH: 3793
/ TYPE: DNA
/ ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.: 1.88e-10 Length: 3793
Score: 145.00 Matches: 26
Percent Similarity: 87.88% Conservative: 3
Best Local Similarity: 78.79% Mismatches: 4
```

```
Query Match: 44.48% Indels: 0
DB: 26 Gaps: 0

US-09-774-954-3 (1-61) x US-11-097-143-18082 (1-3793)
QY 19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
DB 584 GGACGCTTTGGCAACAGCGCGACCACTTCCTGGGATCATTTGGCCTTCGCCAAGGCGCTT 525
QY 39 ValArgThrLeuAlaValProTfPileGluTyrGln 51
DB 524 AATCGCACCTGATCTCCGCCCGTGGTGGATATCGT 486

RESULT 13
/ Sequence 602444, Application US/09925065A
/ Publication No. US20050228172A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single
/ TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.135
/ CURRENT APPLICATION NUMBER: US/09/925,065A
/ CURRENT FILING DATE: 2001-08-08
/ PRIOR APPLICATION NUMBER: US 60/243,096
/ PRIOR FILING DATE: 2000-10-24
/ PRIOR APPLICATION NUMBER: US 60/252,147
/ PRIOR FILING DATE: 2000-11-20
/ PRIOR APPLICATION NUMBER: US 60/250,092
/ PRIOR FILING DATE: 2000-11-30
/ PRIOR APPLICATION NUMBER: US 60/261,766
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/289,846
/ NUMBER OF SEQ ID NOS: 957086
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 602444
/ LENGTH: 521
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-925-065A-602444

Alignment Scores:
Pred. No.: 0.281 Length: 521
Score: 76.00 Matches: 14
Percent Similarity: 83.33% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 3
Query Match: 23.31% Indels: 0
DB: 13 Gaps: 0

US-09-774-954-3 (1-61) x US-09-925-065A-602444 (1-521)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***pro***MetGlyArg 20
DB 280 GCGGCGCTCTCGGACCGCGCGGTACTCTCTACTGCCCCCTGATGGGTAA 227

RESULT 14
US-10-820-474A-236
/ Sequence 236, Application US/10820474A
/ Publication No. US20050155089A1
/ GENERAL INFORMATION:
/ APPLICANT: LAL, PREETI
/ APPLICANT: TANG, Y. TOM
/ APPLICANT: GORGONE, GINA A.
/ APPLICANT: CORLEY, NEIL C.
/ APPLICANT: GUEGLER, KARL J.
/ APPLICANT: BAUGHN, MARIAH R.
/ APPLICANT: AKERBLOM, INGRID E.
/ APPLICANT: AU-YOUNG, JANICE
/ APPLICANT: YUE, HENRY
/ APPLICANT: PATTERSON, CHANDRA
/ APPLICANT: REDDY, ROOPA
/ APPLICANT: HILLMAN, JENNIFER L.
```



```

; APPLICANT: BANDMAN, OLGA
; TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
; FILE REFERENCE: 039386-1568
; CURRENT APPLICATION NUMBER: US/10/820,474A
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: 09/720,533
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: PCT/US99/14484
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/090,762
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/094,983
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: 60/102,686
; PRIOR FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-236

Alignment Scores:
Pred. No.: 0.433 Length: 760
Score: 76.00 Matches: 14
Percent Similarity: 83.33% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 3
Query Match: 23.31% Indels: 0
DB: 24 Gaps: 0

US-09-774-954-3 (1-61) x US-10-820-474A-236 (1-760)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 117 GCGGGCTCTCTGGGACCGCGCGGTACTGCTCTACTGCCCTCGATGGGTAAG 170

RESULT 15
US-10-156-761-5001
; Sequence 5001, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5001
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1803)
US-10-156-761-5001

Alignment Scores:
Pred. No.: 3.83 Length: 1803
Score: 72.50 Matches: 23
```

```

Percent Similarity: 40.26% Conservative: 8
Best Local Similarity: 29.87% Mismatches: 17
Query Match: 22.24% Indels: 29
DB: 16 Gaps: 5

US-09-774-954-3 (1-61) x US-10-156-761-5001 (1-1803)

QY 5 SerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGln 24
Db 178 TCCTGGGACATCAAGAACTAT---CTGTGGGCGCCAGGCTCGGCATGTACGGCATCCAG 234
QY 25 AlaAspHisPheLeu-----GlySer 31
Db 235 CGCATCCACCTCCTGGGCAACGTCATGCTGTGGCGGCGCGCGGGTTCGGCGCGGCTCC 294
QY 32 LeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro----- 46
Db 295 CTCAACTACGCG-----AACACCCTTACGTACCCCGCGAAGGCTTCTTCGAC 342
QY 47 -----Trp-----IleGluTyrGlnHisHisLysProProPhe 57
Db 343 GACCCGAGTGGAGGACATCACCGACTGGCAGGAGGAAGTGAAGCGGTAC 393

Search completed: October 26, 2005, 16:00:04
Job time : 287.326 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 08:18:39 ; Search time 53.4846 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGLYLXPMGR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09774954/runat 25102005 105433 6415/app query.fasta_1.917
-DB=Issued Patents NA -CFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.116 @runat 25102005 105433 6415 -NCPUS=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	295	90.5	1100	3	US-08-978-741-16
2	295	90.5	1100	3	US-09-333-729A-16
3	295	90.5	1300	3	US-08-978-741-4
4	295	90.5	1300	3	US-08-333-729A-6
5	295	90.5	1514	3	US-08-978-741-1
6	295	90.5	1514	3	US-09-333-729A-2
7	295	90.5	11284	3	US-08-978-741-5
8	204	62.6	5009	3	US-08-978-741-7
9	204	62.6	5009	3	US-09-333-729A-8
10	177	54.3	1320	4	US-09-270-767-14353
11	76	23.3	19227	4	US-09-949-016-12127
12	76	23.3	19228	4	US-09-949-016-16285

c	13	68.5	21.0	42325	4	US-08-311-731A-131	Sequence 131, Appl
	14	67	20.6	266293	4	US-09-949-016-11934	Sequence 11934, A
	15	66.5	20.4	2214	4	US-09-583-110-2414	Sequence 2414, Ap
	16	66.5	20.4	2292	4	US-09-107-433-1007	Sequence 1007, Ap
	17	66.5	20.4	2352	2	US-08-922-837-1	Sequence 1, Appli
	18	66.5	20.4	2352	3	US-09-351-550-1	Sequence 1, Appli
	19	66.5	20.4	18627	3	US-08-961-527-113	Sequence 113, App
	20	65	19.9	601	4	US-09-949-016-20390	Sequence 20390,
	21	65	19.9	601	4	US-09-949-016-20391	Sequence 17504, A
	22	65	19.8	175637	4	US-09-949-016-17504	Sequence 16089, A
	23	64.5	19.8	175665	4	US-09-949-016-16089	Sequence 14757, A
	24	62.5	19.2	154600	4	US-09-949-016-14757	Sequence 2, Appli
	25	62	19.0	4403765	3	US-09-103-840A-2	Sequence 10, Appl
	26	62	19.0	4411529	3	US-09-103-840A-1	Sequence 9, Appli
	27	61.5	18.9	999	5	PCT-US91-00899-10	Sequence 1, Appli
	28	61.5	18.9	1136	1	US-08-395-800A-9	Sequence 1, Appli
	29	61.5	18.9	1155	1	US-08-434-151-1	Sequence 1, Appli
	30	61.5	18.9	1155	1	US-08-208-889A-1	Sequence 1, Appli
	31	61.5	18.9	1155	2	US-08-433-271-1	Sequence 1, Appli
	32	61.5	18.9	1155	2	US-08-715-259-1	Sequence 1, Appli
	33	61.5	18.9	1174	5	PCT-US95-07554-3	Sequence 3, Appli
	34	61.5	18.9	1199	1	US-08-395-800A-5	Sequence 5, Appli
	35	61.5	18.9	2268	3	US-08-675-773B-4	Sequence 4, Appli
	36	61.5	18.9	3373	1	US-08-273-411-2	Sequence 2, Appli
	37	61.5	18.9	3374	4	US-09-949-016-5482	Sequence 5482, Ap
	38	61.5	18.9	3791	3	US-08-675-773B-3	Sequence 3, Appli
	39	61.5	18.9	8174	1	US-07-914-281-5	Sequence 5, Appli
	40	61.5	18.9	8174	1	US-08-393-246-5	Sequence 5, Appli
	41	61.5	18.9	8174	1	US-08-525-058A-5	Sequence 5, Appli
	42	61.5	18.9	8174	2	US-08-696-731-5	Sequence 5, Appli
	43	61.5	18.9	8174	3	US-09-042-531-5	Sequence 5, Appli
	44	61.5	18.9	8174	5	PCT-US91-00899-3	Sequence 3, Appli
	45	61.5	18.9	8580	4	US-09-949-016-17224	Sequence 17224, A

ALIGNMENTS

RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

Against DNA

```
/
/ LENGTH: 1100 base pairs
/ TYPE: Nucleic Acid
/ STRANDEDNESS: Single
/ TOPOLOGY: Linear
US-08-978-741-16

Alignment Scores:
Pred. No.: 5.56e-36 Length: 1100
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-08-978-741-16 (1-1100)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 7 GCGGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCGTCATGGGGCGCTTTGGG 66
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 67 AACGAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 126
QY 43 AlaValProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
Db 127 GCTGTCCCTCTGGATTGAGTACCAGATCAACAGCCTCTTTTCCACCACTCCAT 183

RESULT 2
US-09-333-729A-16
/ Sequence 16, Application US/09333729A
/ Patent No. 6270987
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Yang
/ TITLE OF INVENTION: O-Fucosyltransferase
/ FILE REFERENCE: P1041PDI-Substitute
/ CURRENT APPLICATION NUMBER: US/09/333,729A
/ PRIOR FILING DATE: 1999-06-15
/ PRIOR APPLICATION NUMBER: US 08/798,741
/ NUMBER OF SEQ ID NOS: 21
/ SEQ ID NO 16
/ LENGTH: 1100
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-09-333-729A-16

Alignment Scores:
Pred. No.: 5.56e-36 Length: 1100
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-09-333-729A-16 (1-1100)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 7 GCGGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCGTCATGGGGCGCTTTGGG 66
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 67 AACGAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 126
QY 43 AlaValProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
Db 127 GCTGTCCCTCTGGATTGAGTACCAGATCAACAGCCTCTTTTCCACCACTCCAT 183

RESULT 3
US-08-978-741-4
/ Sequence 4, Application US/08978741
/ Patent No. 6100076
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Michael W. Spellman
/ TITLE OF INVENTION: O-Fucosyltransferase
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genentech, Inc.
/ STREET: 1 DNA Way
/ CITY: South San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94080
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WinPatIn (Genentech)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/978,741
/ FILING DATE: 26-Nov. 6100076-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/792498
/ FILING DATE: 31
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Svoboda, Craig C
/ REGISTRATION NUMBER: 39,044
/ REFERENCE/DOCKET NUMBER: P1041P1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650/225-1489
/ TELEFAX: 650/952-9881
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1300 base pairs
/ TYPE: Nucleic Acid
/ STRANDEDNESS: Single
/ TOPOLOGY: Linear
US-08-978-741-4

Alignment Scores:
Pred. No.: 6.98e-36 Length: 1300
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservatives: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-08-978-741-4 (1-1300)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 142 GCGGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCGTCATGGGGCGCTTTGGG 201
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 202 AACGAGCCGATCACTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 261
QY 43 AlaValProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
Db 262 GCTGTCCCTCTGGATTGAGTACCAGATCAACAGCCTCTTTTCCACCACTCCAT 318

RESULT 4
US-09-333-729A-6
/ Sequence 6, Application US/09333729A
/ Patent No. 6270987
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Yang
/ TITLE OF INVENTION: O-Fucosyltransferase
/ FILE REFERENCE: P1041PDI-Substitute
/ CURRENT APPLICATION NUMBER: US/09/333,729A
/ PRIOR FILING DATE: 1999-06-15
/ PRIOR APPLICATION NUMBER: US 08/798,741
```

; PRIOR FILING DATE: 1997-11-26
 ; NUMBER OF SEQ ID NOS: 21
 ; SEQ ID NO 6
 ; LENGTH: 1300
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Plasmid insert.
 US-09-333-729A-6

Alignment Scores:
 Pred. No.: 6.98e-36 Length: 1300
 Score: 295.00 Matches: 54
 Percent Similarity: 91.53% Conservatives: 0
 Best Local Similarity: 91.53% Mismatches: 5
 Query Match: 90.49% Indels: 0
 DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-09-333-729A-6 (1-1300)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 Db 142 GCGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCGCTGCAATGGGGCGCTTTGGG 201
 QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
 Db 202 AACAGGCCGATCCTCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 261
 QY 43 AlaValProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHis 61
 Db 262 GCTGTCCCTCTCTGGATTTGAGTACCAAGCTCCTTTTCAACCACTCCAT 318

RESULT 5

US-08-978-741-1
 ; Sequence 1, Application US/08978741
 ; Patent No. 6100076

GENERAL INFORMATION:

; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978.741
 ; FILING DATE: 26-Nov. 6100076-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/792498
 ; FILING DATE: 31

ATTORNEY/AGENT INFORMATION:

; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1041P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1514 base pairs
 ; TYPE: Nucleic Acid
 ; STRANDEDNESS: Single
 ; TOPOLOGY: Linear

US-08-978-741-1

Alignment Scores:

Pred. No.: 8.58e-36 Length: 1514
 Score: 295.00 Matches: 54
 Percent Similarity: 91.53% Conservatives: 0
 Best Local Similarity: 91.53% Mismatches: 5
 Query Match: 90.49% Indels: 0
 DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-08-978-741-1 (1-1514)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 Db 7 GCGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCGCTGCAATGGGGCGCTTTGGG 66
 QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
 Db 67 AACAGGCCGATCCTCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 126
 QY 43 AlaValProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHis 61
 Db 127 GCTGTCCCTCTCTGGATTTGAGTACCAAGCTCCTTTTCAACCACTCCAT 183

RESULT 6

US-09-333-729A-2
 ; Sequence 2, Application US/09333729A
 ; Patent No. 6270987

GENERAL INFORMATION:

; APPLICANT: Wang, Yang
 ; APPLICANT: Spellman, Michael W.
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; FILE REFERENCE: P1041P1D1-Substitute
 ; CURRENT APPLICATION NUMBER: US/09/333,729A
 ; CURRENT FILING DATE: 1999-06-15
 ; PRIOR APPLICATION NUMBER: US 08/798,741
 ; PRIOR FILING DATE: 1997-11-26
 ; NUMBER OF SEQ ID NOS: 21

SEQ ID NO 2

; LENGTH: 1514
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien

US-09-333-729A-2

Alignment Scores:

Pred. No.: 8.58e-36 Length: 1514
 Score: 295.00 Matches: 54
 Percent Similarity: 91.53% Conservatives: 0
 Best Local Similarity: 91.53% Mismatches: 5
 Query Match: 90.49% Indels: 0
 DB: 3 Gaps: 0

US-09-774-954-3 (1-61) x US-09-333-729A-2 (1-1514)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 Db 7 GCGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCGCTGCAATGGGGCGCTTTGGG 66
 QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
 Db 67 AACAGGCCGATCCTCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 126
 QY 43 AlaValProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHis 61
 Db 127 GCTGTCCCTCTCTGGATTTGAGTACCAAGCTCCTTTTCAACCACTCCAT 183

RESULT 7

US-08-978-741-5
 ; Sequence 5, Application US/08978741
 ; Patent No. 6100076

GENERAL INFORMATION:

; APPLICANT: Yang Wang, Michael W. Spellman
 ; TITLE OF INVENTION: O-Fucosyltransferase
 ; NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 1 DNA Way
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,741
 FILING DATE: 26-No. 6100076-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/792498

FILING DATE: 31

ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.

REGISTRATION NUMBER: 39,044

REFERENCE/DOCKET NUMBER: P1041P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-1489

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 11284 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-978-741-5

Alignment Scores:

Pred. No.:	1,328-34	Length:	11284
Score:	235.00	Matches:	54
Percent Similarity:	91.53%	Conservative:	0
Best Local Similarity:	91.53%	Mismatches:	5
Query Match:	90.49%	Indels:	0
DB:	3	Gaps:	0

US-09-774-954-3 (1-61) x US-08-978-741-5 (1-11284)

QY 3 AlaglySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 DB 4242 GCGGGCTCTCGGAGCCGCGGTACCTGCTACTGCGCCCTGCAATGGGCGCTTTGGG 4301

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
 DB 4302 AACGAGCGGATCACTCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 4361

QY 43 AlavalProProTrrPileGluTyrGlnHisLysProProPheThrAsnLeuHis 61
 DB 4362 GCTGTCCCTCTTGATTGAGTACCAAGCTGCTAAAGCTGCTAAACCGTACCTTG 4418

RESULT 8

US-08-978-741-7

; Sequence 7, Application US/08978741

; Patent No. 6100076

; GENERAL INFORMATION:

; APPLICANT: Yang Wang, Michael W. Spellman

; TITLE OF INVENTION: O-Fucosyltransferase

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 1 DNA Way

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/978,741
 ; FILING DATE: 26-No. 6100076-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/792498

FILING DATE: 31

ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.

REGISTRATION NUMBER: 39,044

REFERENCE/DOCKET NUMBER: P1041P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-1489

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 5009 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-08-978-741-7

Alignment Scores:

Pred. No.:	5,71e-21	Length:	5009
Score:	204.00	Matches:	37
Percent Similarity:	94.87%	Conservative:	0
Best Local Similarity:	94.87%	Mismatches:	2
Query Match:	62.58%	Indels:	0
DB:	3	Gaps:	0

US-09-774-954-3 (1-61) x US-08-978-741-7 (1-5009)

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
 DB 2 AACGAGCGGATCACTCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 61

QY 43 AlavalProProTrrPileGluTyrGlnHisLysProProPheThrAsnLeuHis 61
 DB 62 GCTGTCCCTCTTGATTGAGTACCAAGCTGCTAAAGCTGCTAAACCGTACCTTG 118

RESULT 9

US-09-333-729A-8

; Sequence 8, Application US/09333729A

; Patent No. 6270987

; GENERAL INFORMATION:

; APPLICANT: Wang, Yang

; TITLE OF INVENTION: O-Fucosyltransferase

; FILE REFERENCE: P1041P1D1-Substitute

; CURRENT APPLICATION NUMBER: US/09/333,729A

; CURRENT FILING DATE: 1999-06-15

; PRIOR APPLICATION NUMBER: US 08/798,741

; PRIOR FILING DATE: 1997-11-26

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO 8

; LENGTH: 5009

; TYPE: DNA

; ORGANISM: Homo Sapien

US-09-333-729A-8

Alignment Scores:

Pred. No.:	5,71e-21	Length:	5009
Score:	204.00	Matches:	37
Percent Similarity:	94.87%	Conservative:	0
Best Local Similarity:	94.87%	Mismatches:	2
Query Match:	62.58%	Indels:	0
DB:	3	Gaps:	0

US-09-774-954-3 (1-61) x US-09-333-729A-8 (1-5009)

QY 23 AenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeuLeu***ValArgThrLeu 42
Db 2 AACAGCGCATCACCTCTGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 61
QY 43 AlaValProProTrpIleGluTyrGlnHisHisLysPheProPheThrAsnLeuHis 61
Db 62 GCTGTCCCTCTCTGGATGAGTACCAAGCATCACAGCTCTTTTACCAACCACTCCAT 118

RESULT 10

US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14353

Alignment Scores:

Pred. No.: 1.43e-17 Length: 1320
Score: 177.00 Matches: 33
Percent Similarity: 80.00% Conservative: 3
Best Local Similarity: 73.33% Mismatches: 9
Query Match: 54.29% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-3 (1-61) x US-09-270-767-14353 (1-1320)

QY 7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAenGlnAlaAsp 26
Db 75 GATCCCATGGCTACCTCTACCTACTGCTGTATGGAGCGCTTGGCAACGAGCGGAC 134
QY 27 HisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro 46
Db 135 CACTTCTGGATCATTTGGCTTCGCAAGCGCTTAATCGCACCTGTATCTCGCGCG 194
QY 47 TrpIleGluTyrGln 51
Db 195 TGGGTGGAGTATCGT 209

RESULT 11

US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12127
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12127

Alignment Scores:
Pred. No.: 2.55 Length: 19227
Score: 76.00 Matches: 14
Percent Similarity: 83.33% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 3
Query Match: 23.31% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-3 (1-61) x US-09-949-016-12127 (1-19227)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 1712 GCGGGCTCTGGACCCGCGGTACTGCTTACTGCCCTGCAATGGGTAAG 1659

RESULT 12

US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16285
; LENGTH: 19228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16285

Alignment Scores:

Pred. No.: 2.55 Length: 19228
Score: 76.00 Matches: 14
Percent Similarity: 83.33% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 3
Query Match: 23.31% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-3 (1-61) x US-09-949-016-16285 (1-19228)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 1712 GCGGGCTCTGGACCCGCGGTACTGCTTACTGCCCTGCAATGGGTAAG 1659

RESULT 13

US-08-311-731A-131/c
; Sequence 131, Application US/08311731A
; Patent No. 6582266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311.731A

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: C0044/7125

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-3500

INFORMATION FOR SEQ ID NO: 131:

SEQUENCE CHARACTERISTICS:

LENGTH: 42325 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: MYCOBACTERIUM LEPRAE

US-08-311-731A-131

Alignment Scores:

Pred. No.: 109

Score: 42325

Percent Similarity: 23

Best Local Similarity: 6

Mismatches: 42.65%

Query Match: 33.82%

Indels: 25

Gaps: 4

US-09-774-954-3 (1-61) x US-08-311-731A-131 (1-42325)

QY 5 SerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGln 24

Db 30298 TCTGGGATCTCGTAATTC---CTTGGCGCCGGAAGCTGGGTGCTACGGCATCCAA 30242

QY 25 AlaAspHisPheLeu-----GlySer 31

Db 30241 CGCATTCACCTCTCGCAACGTCATGATATTCGGCGCGCGGAGTGGGAGCGGGTTCG 30182

QY 32 LeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrpIleGluTyrGln 51

Db 30181 TTGAACCTACGCG-----AATACGTTGTACGTTCCGCCG----- 30149

QY 52 HisHisLysProProPheThrAsn 59

Db 30148 -----GAGCGGTCTTTGCCAC 30131

RESULT 14

US-09-949-016-11934

; Sequence 11934, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; PRIOR FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11934

; LENGTH: 266293
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11934

Alignment Scores:

Pred. No.: 226293

Score: 19

Percent Similarity: 67.00

Best Local Similarity: 45.45%

Query Match: 28.79%

Indels: 22

Gaps: 3

US-09-774-954-3 (1-61) x US-09-949-016-11934 (1-266293)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20

Db 236166 CGACACCTGCTTCCTGGTTCATGATGGGCATCTTCTACTCTTACATCGGACGA 236225

QY 21 PheGlyAsnGlnAlaAspHisPheLeu-----GlySerLeuAlaPheAlaLys--- 36

Db 236226 AAGGGAGGAGAGAGCTCTCTGTGGTCTCTTTTATGAGGGGCTTACTCTCATGCAAGAC 236285

QY 37 Leu***ValArgThrLeuAlaValProProTrp----- 47

Db 236286 CTCACCTTCATGACTAGTCATCCCGAGTGGTCCACCTCTTAATACCATCCTCAG 236345

QY 48 ---IleGluTyrGlnHis 52

Db 236346 GAGTAGAATTTCAACAT 236363

RESULT 15

US-09-583-110-2414

; Sequence 2414, Application US/09583110

; Patent No. 6699703

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.

; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus

; Pneumoniae for Diagnostics and Therapeutics

; FILE REFERENCE: PATH00-07A

; CURRENT APPLICATION NUMBER: US/09/583,110

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30

; PRIOR APPLICATION NUMBER: US 60/085,131

; PRIOR FILING DATE: 1998-05-12

; PRIOR APPLICATION NUMBER: US 60/051,553

; PRIOR FILING DATE: 1997-07-02

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 2414

; LENGTH: 2214

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-583-110-2414

Alignment Scores:

Pred. No.: 4.01

Score: 2214

Percent Similarity: 66.50

Best Local Similarity: 48.84%

Query Match: 37.21%

Indels: 19

Gaps: 1

US-09-774-954-3 (1-61) x US-09-583-110-2414 (1-2214)

QY 8 LeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAspHis 27

Db 418 TTGATCGGGTTCGCTCTTTATATTCACACAGCCTTCTCTTTTCAAAATATCGGAACCTTAC 477

QY 28 PheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrp 47

Db 478 TTATATGGTTCATCTTG-----ATTTAGTGGGTTCCTCTCTAGTCAGCCCTTGG 528

Qy 48 IleGluTyr 50
Db 529 TCTGTTTAC 537

Search completed: October 26, 2005, 15:25:31
Job time : 106.485 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 03:15:37 ; Search time 170.85 Seconds
(without alignments)
2113.573 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGLYLXPXMR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat_25102005_105431_6380/app_query.fasta_1.917
-DB=N_Geneseq -QFWT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USBR=US09774954 @CGN 1.1 703 @runat_25102005_105431_6380 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1980s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	295	90.5	1514	2	AAV65632	Human hea
2	295	90.5	5218	13	ACN40746	Tumour-as
3	295	90.5	5266	10	ADF81754	Leukaemia
4	295	90.5	11284	2	AAV65633	Plasmid c
5	232.5	71.3	5230	4	AAK51510	Human pol

6	220.5	67.6	4850	4	AAK52494	Human pol
7	204	62.6	5009	2	AAV65634	First Eco
8	177	54.3	1209	4	ABL13947	Drosophil
9	145	44.5	3264	4	ABL13946	Drosophil
10	145	44.5	3793	4	ABL13894	Drosophil
11	76	23.3	760	3	AAZ98210	Human sig
12	69.5	21.3	594	8	ACA37736	Prokaryot
13	68.5	21.0	42325	10	ADB74382	Mycobacte
14	67.5	20.7	2720	10	ADB53234	Primary r
15	66.5	20.4	1660	2	AAZ96329	S. pneumo
16	66.5	20.4	2223	10	ABX06561	S. pneumo
17	66.5	20.4	2292	13	ADR92372	Novel S.
18	66.5	20.4	2304	3	AAA05535	Streptoco
19	66.5	20.4	2304	4	AAZ55608	Streptoco
20	66.5	20.4	2304	4	AAZ55863	Streptoco
21	66.5	20.4	2304	8	ACA49857	Prokaryot
22	66.5	20.4	2351	2	AAZ16136	DNA encod
23	66.5	20.4	18627	2	AAV52246	Streptoco
24	66.5	20.4	110000	10	ABS56454_07	Continuatio
25	65	19.9	406	10	ADB50839	Primary r
26	65	19.9	478	10	ABT40319	Toxicity
27	65	19.9	520	10	ADB56829	Toxicity
28	65	19.9	97081	12	ADQ97980	Human can
29	65	19.9	110000	5	AAI61373_0	Continuatio
30	65	19.9	110000	5	AAI61373_1	Continuatio
31	64.5	19.8	575	12	ADJ75816	Marker ge
32	64.5	19.8	2155	10	ADI23671	Mouse LPD
33	64	19.6	1018	8	ACA49002	Prokaryot
34	64	19.6	1580	8	ADA69987	Rice gene
35	64	19.6	4184	4	AAK83286	Human imm
36	63.5	19.5	510	12	ACH76865	Human gen
37	63.5	19.5	5983	4	AAK84657	Human imm
38	63	19.3	577	6	ABL93098	Rat metas
39	63	19.3	577	12	ADN07889	Human mam
40	63	19.3	8640	4	ABL11790	Drosophil
41	62.5	19.2	1314	8	ACA36930	Prokaryot
42	62.5	19.2	3315	11	ADO58598	Porcine a
43	62.5	19.2	73507	11	ACN44612	Mouse gen
44	62.5	19.2	110000	6	ABA03041_03	Continuatio
45	62	19.0	8455	13	ABD33038	Mouse can

ALIGNMENTS

RESULT 1
AAV65632 standard; DNA; 1514 BP.
XX AC AAV65632;
XX 16-DEC-1998 (first entry)
XX Human heart O-fucosyltransferase encoding DNA.
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
XX O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX misc_feature 1..1100
XX /*tag= b
XX /note= "this actively expressed O-fucosyltransferase
XX sequence is claimed for in claim 9"
XX CDS 1..1098
XX /*tag= a
XX /product= "human heart O-fucosyltransferase"

Against DNA

```
XX 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX (GETH ) GENENTECH INC.
PA Wang Y, Spellman MW;
PI WPI; 1998-437477/37.
XX P-FSDS; AAW80571.
DR Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX Claim 9; Fig 12A; 90pp; English.
XX This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EGF domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC probes based on EGF domain polypeptide are used to detect gene
CC amplification and expression. The expression can also be determined at
CC the protein level using antibodies specific for O-fucosyltransferase
XX SQ Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 5,878-33 Length: 1514
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-3 (1-61) x AAV65632 (1-1514)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
DB 7 GCGGGCTCTCTGGGACCCGCGGTACTGCTCTACTGCCCTGTCATGGGCGCTTTGGG 66
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
DB 67 AACCAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 126
QY 43 AlaValProProTrpIleGlyTyrGlnHisLeuProProPheThrAsnLeuHis 61
DB 127 GCTGTCCCTCTTGGATTGAGTACCAGCATCACAGCTCTCTTCAACCACTCCAT 183

RESULT 2
ACN40746
ID ACN40746 standard; cDNA; 5218 BP.
AC ACN40746;
XX 18-NOV-2004 (first entry)
XX Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX Homo sapiens.
XX WO2004030615-A2.
XX
```

```
PD 15-APR-2004.
XX 29-SEP-2003; 2003WO-US028547.
XX 02-OCT-2002; 2002US-0414971P.
XX (GETH ) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
XX WPI; 2004-347921/32.
XX P-FSDS; ABM82223.
XX New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX Claim 1; SEQ ID NO 5725; 7273pp; English.
XX The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX SQ Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,858-32 Length: 5218
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 13 Gaps: 0

US-09-774-954-3 (1-61) x ACN40746 (1-5218)
QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
DB 125 GCGGGCTCTCTGGGACCCGCGGTACTGCTCTACTGCCCTGTCATGGGCGCTTTGGG 184
QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
DB 185 AACCAGGCGGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 244
QY 43 AlaValProProTrpIleGlyTyrGlnHisLeuProProPheThrAsnLeuHis 61
DB 245 GCTGTCCCTCTTGGATTGAGTACCAGCATCACAGCTCTCTTCAACCACTCCAT 301

RESULT 3
ADF81754
ID ADF81754 standard; DNA; 5266 BP.
XX ADF81754;
XX 26-FEB-2004 (first entry)
XX
```


RESULT 5
 ID AAK51510 standard; cDNA, 5230 BP.
 AC AAK51510;
 DT 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 55.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 OS WO200157190-A2.
 PN 09-AUG-2001.
 PD 05-FEB-2001; 2001WO-US004098.
 PF 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663551.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 XX Ma Y, Zhao QA, Wang B, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX P-PSDB; AAM78377.
 XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
 XX in diagnosis and gene therapy.
 XX Claim 1; Page 638-642; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 XX cytokine, cell proliferation or cell differentiation or which may induce
 XX production of other cytokines in other cell populations. The
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
 XX peptide therapy. The polypeptides have various cytokine-like activities,
 XX e.g. stem cell growth factor activity, haematopoiesis regulating
 XX activity, tissue growth factor activity, immunomodulatory activity and
 XX activity/inhibit activity and may be useful in the diagnosis and/or
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
 XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 XX (AAK52592) and 3666 (AAM80020) are omitted as the relevant pages from the
 XX sequence listing were missing at the time of publication
 XX Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred No.: 4.11e-23 Length: 5230
 Score: 232.50 Matches: 44
 Percent Similarity: 80.00% Conservative: 7
 Best Local Similarity: 73.33% Mismatches: 4
 Query Match: 71.32% Indels: 5
 DB: 4 Gaps: 1
 US-09-774-954-3 (1-61) x AAK51510 (1-5230)
 7 AsnLeuAlaGlyTyrLeuLeuTyr***Pro***Met-----GlyArgPhe 21

us-09-774-954-3.rng

Thu Oct 27 11:17:25 2005

Best Local Similarity:	73.33%	Mismatches:	9
Query Match:	54.29%	Indels:	0
DB:	4	Gaps:	0
US-09-774-954-3 (1-61) x ABL13947 (1-1209)			
QY	7	AspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyValGlnAlaAsp	26
DB	82	GATCCCAATGGTACTCCTACTGCTGCTGATGGACGCTTGGACACCGACGAC	141
QY	27	HisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro	46
DB	142	CACCTCTGGGATCATTCGCTTCGCCAGGCCCTTAATCGCACCTGATCTCGCGCG	201
QY	47	TpPileGlyTyrGln 51	
DB	202	TGGGTGGAGTATCGT 216	
RESULT 9			
ABL13946/c			
ID	ABL13946	standard; cDNA; 3264 BP.	
XX			
AC	ABL13946;		
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 36320.	
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB69843.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions.		
XX			
PS	Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA		
CC	sequences (ABU161840-ABL16175) and the encoded proteins (AB857737-		
CC	AB872072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SQ	Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1,49e-10	Length:	3264
Score:	145.00	Matches:	26
Percent Similarity:	87.88%	Conservative:	3
Best Local Similarity:	78.79%	Mismatches:	4
Query Match:	78.79%	Indels:	0
DB:	4	Gaps:	0
US-09-774-954-3 (1-61) x ABL13894 (1-3793)			
Query Match: 44.48%			
DB: 4			
US-09-774-954-3 (1-61) x ABL13946 (1-3264)			
QY	19	GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***	38
DB	2092	GGACGCTTTGGCAACAGCGCCACCTCTCTGGGATCATTCGCTTCGCCAGCGCCTT	2033
QY	39	ValArgThrLeuAlaValProProTpPileGlyTyrGln 51	
DB	2032	AATCGCACCTGATCTCTCGCGCGCTGGGTGGAGTATCGT 1994	
RESULT 10			
ABL13894/c			
ID	ABL13894	standard; cDNA; 3793 BP.	
XX			
AC	ABL13894;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 36164.	
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
KW	pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB69791.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions.		
XX			
PS	Claim 1; SEQ ID NO 36164; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA		
CC	sequences (ABU161840-ABL16175) and the encoded proteins (AB857737-		
CC	AB872072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SQ	Sequence 3793 BP; 1020 A; 853 C; 950 G; 970 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	1.81e-10	Length:	3793
Score:	145.00	Matches:	26
Percent Similarity:	87.88%	Conservative:	3
Best Local Similarity:	78.79%	Mismatches:	4
Query Match:	44.48%	Indels:	0
DB:	4	Gaps:	0

QY 19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
 Db 584 GGACGGTTTGGCAACACAGCCGACCACTCTCTGGGATCATTTGGCTTGGCCAGGCGCTT 525
 QY 39 ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
 Db 524 AATCGCACCTGATCTCGCCGCGTGGTGGAGTATCGT 486

RESULT 11
 AAZ98210
 ID AAZ98210 standard; cDNA; 760 BP.
 AC AAZ98210;
 XX
 DT 11-MAY-2000 (first entry)
 XX
 DE Human signal peptide containing protein HSPP-102 cDNA SEQ ID NO:236.
 XX
 KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;
 KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;
 KW muscular dystrophy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200000610-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US014484.
 XX
 PR 26-JUN-1998; 98US-0090762P.
 PR 31-JUL-1998; 98US-0094983P.
 PR 01-OCT-1998; 98US-0102686P.
 PR 11-DEC-1998; 98US-0112129P.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
 PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
 PI Bandman O;
 XX
 DR WPI: 2000-160673/14.
 DR P-PSDB; AAY87325.
 XX
 PT New human signal peptide-containing proteins useful in treatment,
 PT prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
 PT disease.
 XX
 PS Claim 9; Page 307; 327pp; English.
 XX
 CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
 CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
 CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can be
 CC used in gene therapy. HSPPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSPP. Antagonists of
 CC HSPP are used to treat or prevent disorders associated with increased
 CC activity or function of HSPP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
 CC nucleic acids can be used for the recombinant production of HSPP, for
 CC detecting HSPP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming

CC or ribozyme therapeutics, for detecting related sequences or genetic
 CC variations, and for chromosomal mapping. HSPP are also used to raise
 CC specific antibodies (Ab) and to screen for agonists and antagonists
 CC (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
 CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
 CC competitive drug screens, and for purification of HSPP from natural
 CC sources
 XX
 SQ Sequence 760 BP; 171 A; 206 C; 220 G; 163 T; 0 U; 0 Other;

Alignment Scores: 0.3 Length: 760
 Pred. No.: 76.00 Matches: 14
 Score: 83.33% Conservative: 1
 Percent Similarity: 77.78% Mismatches: 3
 Best Local Similarity: 23.31% Indels: 0
 Query Match: 3 Gaps: 0
 DB:

US-09-774-954-3 (1-61) x AAZ98210 (1-760)
 QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
 Db 117 GCGGGCTCTGGGACCCGCCGGTTACTGCTTACTGCCCTGCATGGGTAG 170

RESULT 12
 ACA37736
 ID ACA37736 standard; DNA; 594 BP.
 XX
 AC ACA37736;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #19393.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Mycobacterium avium.
 XX
 PN WO20027183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI: 2003-029926/02.
 DR P-PSDB; ABU33866.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 25606; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding

PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX WPI; 2003-731472/69.
XX
XX Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
XX Claim 44; SEQ ID NO 3776; 874pp; English.
XX
XX The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
XX Sequence 2720 BP; 624 A; 719 C; 725 G; 652 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 26.9 Length: 2720
Score: 67.50 Matches: 26
Percent Similarity: 34.74% Conservative: 7
Best Local Similarity: 27.37% Mismatches: 19
Query Match: 10 Indels: 43
DB: 10 Gaps: 5
US-09-774-954-3 (1-61) x ADB53234 (1-2720)
QY 2 LeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPhe 21
Db 713 CTCTCTGGAGCTTGACAAATC-----ACCCCGGA---GGCAGGTTT 751
QY 22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg--- 40
Db 752 GGTAAACAGATGGGACAGATGACATGCTGCGCCCTAGCCAGCTCAATGGCGCGG 811
QY 41 -----ThrLeuAla----- 43
Db 812 GCTTTTCATCCAGCTGAGATGATACACCCCTGGCCCTGTGTCCGAATCTCCCTGCCA 871
QY 44 -----ValProProTrp----- 47

Db 872 GTGCTGGACCTCGAGGTGGACAGCTCTCAGCGCTTGGCAGCACTTAGTCTTACATGACTGG 931
QY 48 -----IlleGluTyrGlnHisLysProProPheThrAsnLeu 60
Db 932 ATGTCAGAGGAGTACTCCACCTGGAGGAGCCCGTTTCTCAAGCTG 976
RESULT 15
AAZ96329
ID AAZ96329 standard; DNA; 1660 BP.
XX
XX AAZ96329;
XX
XX 10-APR-2000 (first entry)
XX
XX S. pneumoniae derived DNA from ORF #157.
XX Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
XX Streptococcus pneumoniae.
XX
XX WO9806734-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US014436.
XX
XX 16-AUG-1996; 96US-0024022P.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
XX WPI; 1998-159452/14.
DR P-PSDB; AAY85984.
XX
XX Streptococcus pneumoniae proteins and related DNA - useful for screening
PT compounds for antibacterial activity.
XX
XX Claim 4; Page 191-192; 640pp; English.
XX
XX This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC AAY85792-Y86182). The DNA, vectors and host cells described in the method
CC of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful for
CC inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease
XX
XX Sequence 1660 BP; 467 A; 332 C; 379 G; 479 T; 0 U; 3 Other;
SQ
Alignment Scores:
Pred. No.: 20.1 Length: 1660
Score: 66.50 Matches: 16
Percent Similarity: 48.84% Conservative: 5
Best Local Similarity: 37.21% Mismatches: 19
Query Match: 20.40% Indels: 3
DB: 2 Gaps: 1
US-09-774-954-3 (1-61) x AAZ96329 (1-1660)
QY 8 LeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAspHis 27
Db 1077 TTGATCGGGTGGCTCTTTTATTTCCACAGCCCTTCTCTTTCAATATCGGAATTAC 1136

```
Qy      28 pheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTirp 47
Db      1137 TTTATTTGGTTCTATCTTG-----ATTTTAGTGGGTTCTCTCTAGTCAGCCCTTGG 1187
Qy      48 IleGluTyr 50
Db      1188 TCTGTTTAC 1196
```

Search completed: October 26, 2005, 09:51:41
Job time : 185.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:08:04 ; Search time 1382.21 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLGALYLLXPXMG.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPO.spool_h/US09774954/runat_25102005_105432_6389/app_query.fasta_1.917
-DB=GenEmbl -OFMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN_1_1_4968 @runat_25102005_105432_6389 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_ey.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	305	93.6	1302	10 AF375885	AF375885 Mus muscu
2	305	93.6	1529	10 BC046295	BC046295 Mus muscu
3	302	92.6	1188	10 AJ781499	AJ781499 Rattus no
4	295	90.5	1100	6 AR163457	AR163457 Sequence

5	295	90.5	1100	6	BD103196	BD103196 O-fucosyl
6	295	90.5	1167	6	CQ727777	CQ727777 Sequence
7	295	90.5	1167	9	AJ781500	AJ781500 Pan trogl
8	295	90.5	1300	6	AR163452	AR163452 Sequence
9	295	90.5	1300	6	BD103189	BD103189 O-fucosyl
10	295	90.5	1514	6	AR163451	AR163451 Sequence
11	295	90.5	1514	6	BD103188	BD103188 O-fucosyl
12	295	90.5	1524	9	BC000582	BC000582 Homo sapi
13	295	90.5	4560	9	HSB803330	HSB803330 Homo sapi
14	295	90.5	5189	9	D80002	D80002 Homo sapien
15	295	90.5	5249	9	AF375884	AF375884 Homo sapi
16	295	90.5	5266	6	AX780153	AX780153 Sequence
17	295	90.5	11284	6	BD103190	BD103190 O-fucosyl
18	288	88.3	1056	4	AY344581	AY344581 Bos tauru
19	288	88.3	1176	4	AY344580	AY344580 Bos tauru
20	286	87.7	1134	4	SSC567917	SSC567917 Sus scrof
21	273	83.7	1176	4	AJ781503	AJ781503 Bos tauru
22	272	83.4	920	5	AJ719585	AJ719585 Gallus ga
23	272	83.4	1732	5	GGAS35754	GGAS35754 Gallus ga
24	272	83.4	2546	5	AJ720352	AJ720352 Gallus ga
25	250	76.7	1502	5	BC082519	BC082519 Xenopus t
26	250	76.7	1538	5	AJ781498	AJ781498 Silurana
27	249	76.4	1549	5	DRES79536	DRES79536 Danio rer
28	247	75.8	1155	5	AJ606070	AJ606070 Fugu rubr
29	247	75.8	1164	5	AJ781504	AJ781504 Tetraodon
30	234	71.8	1505	5	XLA514425	XLA514425 Xenopus l
31	220	67.5	145414	9	HSA392M18	HSA392M18 Human DNA
32	220	67.5	200329	10	AC078911	AC078911 Mus muscu
33	220	67.5	208614	10	AL807380	AL807380 Mouse DNA
34	220	67.5	240336	2	AC094844	AC094844 Rattus no
35	220	67.5	240510	2	AC134371	AC134371 Rattus no
36	204	62.6	5009	6	AR163453	AR163453 Sequence
37	204	62.6	5009	6	BD103191	BD103191 O-fucosyl
38	177	54.3	1209	3	AJ831490	AJ831490 Drosophil
39	177	54.3	1209	6	CQ590403	CQ590403 Sequence
40	177	54.3	1295	3	AY118651	AY118651 Drosophil
41	177	54.3	1320	6	AR509393	AR509393 Sequence
42	177	54.3	1579	3	AB093572	AB093572 Drosophil
43	170	52.1	1146	3	AJ781502	AJ781502 Ciona sav
44	164	50.3	1209	3	AJ781501	AJ781501 Drosophil
45	162	49.7	1841	3	AK112708	AK112708 Ciona int

ALIGNMENTS

RESULT 1	AF375885	1302 bp	mRNA	linear	ROD 23-OCT-2001
LOCUS	Mus musculus protein o-fucosyltransferase (Pofut1)	mrna			complete
DEFINITION	cds.				
ACCESSION	AF375885				
VERSION	AF375885.1	GI:15825115			
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 1302)				
AUTHORS	Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P. and Haltiwanger,R.S.				
TITLE	Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase				
JOURNAL	J. Biol. Chem. 276 (43), 40338-40345 (2001)				
MEDLINE	21523965				
PUBMED	11524432				
REFERENCE	2 (bases 1 to 1302)				
AUTHORS	Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W. and Haltiwanger,R.S.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave., New York, NY 10461, USA				
FEATURES	Location/Qualifiers				

```
source
1. .1302
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129/SvJ"
/db_xref="taxon:10090"
1. .1302
/genes="Pofut1"
/genes="Pofut1"
25. .1206
/genes="O-FucT-1"
/codon_start=1
/product="protein o-fucosyltransferase"
/db_xref="GI:15825116"
/translation="MGAAWAPPHLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
CMGRGNQADHFLGSLAFKLNRTLAVPMTIEYOHKHPFTNLHVSYQKIFKLEPLQ
AYHRVVSLEDFMENLAPSHWPPEKRYAYCFVAAQSPDKCTCPMKEGPNFGPFWDQF
HVSFNKSLFTGTISFSASYKEQWTFPAKHPVLALFGAPQFPVLEHRELQKYMV
WSDENVRTEALISAHLVRYVGIHLRIGSDWKACAMKDXGTAGSHFASPCQGVYS
RSTATPLMTWMLCPDLKEIQRAVTLWRLNARSVYIATDSYVSVEIQOLFQDKVRV
VSLKPEVAQIDLYLQADHFIQNCVSVFPAFKRERDLHGQSFPFGMDRPSLRDE
P"

ALIGNMENT SCORES:
Pred No.: 1-2e-33 Length: 1302
Score: 305.00 Matches: 56
Percent Similarity: 91.80% Conservative: 0
Best Local Similarity: 91.80% Mismatches: 5
Query Match: 93.36% Indels: 0
DB: 10 Gaps: 0

US-09-774-954-3 (1-61) x AF375885 (1-1302)
QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
Db 109 CGGTCAAGCGGCTCTCGGACCTCTGCGGTACTCTCTCTGCTGCTGCTGCGGCGC 168
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
Db 169 TTTGGGAACAGCGCTGATCACTCTCTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCCG 228
QY 41 ThrLeuAlaValProProTrrpIleGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 229 ACCTTGGCTGTACCTCCATGGATTGAATCAACATCAACAGCTCTCTTTCACCAACCTC 288
QY 61 His 61
Db 289 CAT 291

RESULT 2
BC046295 1529 bp mRNA linear ROD 30-JUN-2004
LOCUS Mus musculus protein o-fucosyltransferase 1, mRNA (cDNA clone
DEFINITION MGC:54837 IMAGE:6485896), complete cds.
ACCESSION BC046295
VERSION BC046295.1 GI:28279489
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1529)
REFERENCE
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McSwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smalios,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1529)
Strausberg,R.
Direct submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Contact: nisc_mgc@hgrl.nih.gov
Web site: http://www.nisc.nih.gov/
Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 100 Row: m Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 17978277.
FEATURES
Location/Qualifiers
1. .1529
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="MGC:54837 IMAGE:6485896"
/tissue_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI CGAP Mam2"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. .1529
/genes="Pofut1"
/genes="Pofut1"
/notes="synonym: O-FucT-1"
/db_xref="LocusID:140484"
/db_xref="MGI:2153207"
56. .1237
/genes="Pofut1"
/codon_start=1
/product="protein O-fucosyltransferase 1"
/protein_id="AAH46295.1"
/db_xref="GI:28279490"
/db_xref="LocusID:140484"
/db_xref="MGI:2153207"
/translation="MGAAWAPPHLLRLASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
CMGRGNQADHFLGSLAFKLNRTLAVPMTIEYOHKHPFTNLHVSYQKIFKLEPLQ
AYHRVVSLEDFMENLAPSHWPPEKRYAYCFVAAQSPDKCTCPMKEGPNFGPFWDQF
HVSFNKSLFTGTISFSASYKEQWTFPAKHPVLALFGAPQFPVLEHRELQKYMV
WSDENVRTEALISAHLVRYVGIHLRIGSDWKACAMKDXGTAGSHFASPCQGVYS
RSTATPLMTWMLCPDLKEIQRAVTLWRLNARSVYIATDSYVSVEIQOLFQDKVRV
VSLKPEVAQIDLYLQADHFIQNCVSVFPAFKRERDLHGQSFPFGMDRPSLRDE
P"
gene
CDS
```

WSDVMRTGTEALISAHLVPPYGIHLRIGSDWKNACMLKDGDTAGSHFASPOCVGYS
RSTATPLTMTMCLPDLKEIQRATLWVRALNARSVYIATDSSEYSEIQOLFQDKVRV
VSUKPEVAQIDLYLGQADHFGNVCSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
F"

ORIGIN

Alignment Scores: 1.43e-33 Length: 1529
Pred. No.: 305.00 Matches: 56
Score: 91.80% Conservatives: 0
Percent Similarity: 91.80% Mismatches: 5
Best Local Similarity: 93.56% Indels: 0
Query Match: 10 Gaps: 0
DB: 10

US-09-774-954-3 (1-61) x BC046295 (1-1529)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
|||
Db 140 CGGTACAGCGGCTCTGGGACCTGGCGGTACCTGCTCTACTGCTCCGTCATGGCGCGC 199
|||
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
|||
Db 200 TTTGGGAACAGGCTGATCATCTCTTGGGCTCCCTGCGCATTTGCGAAGCTGCTGAACCGC 259
|||
QY 41 ThrLeuAlaValProTrpPheGluTyrGlnHisLysProProPheThrAsnLeu 60
|||
Db 260 ACCTTGCTGTACTCCATGGATGTAATACCAATCACAAGCTCTTTTCCACCACTC 319
|||
QY 61 His 61
|||
Db 320 CAT 322

RESULT 3
AJ781499 1188 bp mRNA linear ROD 07-JUL-2004
LOCUS Rattus norvegicus mRNA for protein-O-fucosyltransferase 1 (fuc12
DEFINITION

ACCESSION AJ781499
VERSION AJ781499.1 GI:50057067
KEYWORDS fuc12 gene; put1 gene; protein-O-fucosyltransferase 1.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 Martinez-Duncker, I., Mollicone, R., Candellier, J.J., Breton, C. and Oriol, R.
AUTHORS

TITLE A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs
JOURNAL Glycobiology 13 (12), 1C-5C (2003)
PUBMED 12966037

REFERENCE 2 Martinez-Duncker, I., Oriol, R. and Mollicone, R.
AUTHORS

TITLE Phylogeny of fucosyltransferases
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1188)
AUTHORS Oriol, R.

TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE
FEATURES

source
1. .1188
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Wistar"
/db_xref="taxon:10116"
1. .1188
/gene="fuc12"
1. .1188
/gene="fuc12"
/function="adds fucose to Ser or Thr"

/note="pofut1"
/codon_start=1
/product="protein-O-fucosyltransferase 1"
/protein_id="CAH03711.1"
/db_xref="GI:50057068"

/translation="MGAAAPASHLLLRASFLLLLLLLPLPLRGRAGGSWDLAGYLLY
CPGMRFGNQADHFLGSLAFKLLRNTLAVPFIYQHHKPPFTNLHVSYQKFKLEP
LQAHVRVLSLEDFMEKLPFFWPPKRVAYCFEVAARSPDKKTCMKEGNPFPGWPD
QPHVFNKSELFTGISFSAKYEQWIPPPKEHPVLPALPGAPAPFVLEHRELOKY
MWSDEMVRTGSAQISTHLVRPVGILHRLIGSDWKNACMLKDGDTAGSHFASPOCVG
YSRSTATPLTMTMCLPDLKEIQRATLWVRALNARSVYIATDSSEYSEIQOLFQDKVRV
KVSUKPEVAQIDLYLGQADHFGNVCSSFTAFVKRERDLHGRPSFFGMDRPSQLR
DEF"

ORIGIN

Alignment Scores: 2.91e-33 Length: 1188
Pred. No.: 302.00 Matches: 55
Score: 90.16% Conservatives: 0
Percent Similarity: 90.16% Mismatches: 6
Best Local Similarity: 92.64% Indels: 0
Query Match: 10 Gaps: 0
DB: 10

US-09-774-954-3 (1-61) x AJ781499 (1-1188)

QY 1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
|||
Db 91 CGGCGCGGCGGCTCTGGGATCTGGCGGTACTGCTCTACTGCTCCGTCATGGCGCGC 150
|||
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
|||
Db 151 TTTGGGAACAGGCTGATCATCTCTTGGGCTCTCTGGCATTTTCAAAGCTGCTGAACCGC 210
|||
QY 41 ThrLeuAlaValProTrpPheGluTyrGlnHisLysProProPheThrAsnLeu 60
|||
Db 211 ACCTTGCTGTACTCCATGGATGTAATACCAATCACAAGCTCTCCCTTCCACCACTC 270
|||

QY 61 His 61
|||
Db 271 CAT 273

RESULT 4
AR163457 1100 bp DNA linear PAT 17-OCT-2001
LOCUS AR163457
DEFINITION Sequence 16 from patent US 6270987.
ACCESSION AR163457
VERSION AR163457.1 GI:16234062
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 16 07-AUG-2001;
FEATURES Location/Qualifiers
source
1. .1100
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores: 2.69e-32 Length: 1100
Pred. No.: 295.00 Matches: 54
Score: 91.53% Conservatives: 0
Percent Similarity: 91.53% Mismatches: 5
Best Local Similarity: 90.49% Indels: 0
Query Match: 6 Gaps: 0
DB: 6

US-09-774-954-3 (1-61) x AR163457 (1-1100)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
|||

Db 7 GCGGGCTCTGGACCGCGGTTACTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 66
Qy 23 AenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 67 AACAGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 126
Qy 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db 127 GCTGTCCCTCTGGATTGAGTACCAAGCATCAAGCCCTCTTTTACCAACCTCCAT 183
RESULT 5
LOCUS BD103196 1100 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103196
VERSION BD103196.1 GI:22648770
KEYWORDS JP 2001527389-A/9.
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1100)
AUTHORS Wang,Y and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 9 25-DEC-2001;
GENENTECH INC
COMMENT OS Unidentified
PN JP 2001527389-A/9
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
PC C12N15/54, C12N9/10, C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FT Key Location/Qualifiers
FT source 1. .1100
FT /organism='Unidentified'.
FEATURES
source 1. .1100
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'
ORIGIN
Alignment Scores: 2.59e-32 Length: 1100
Pred. No.: 295.00 Matches: 54
Score: 295.00
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 6 Gaps:
US-09-774-954-3 (1-61) x BD103196 (1-1100)
Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 7 GCGGGCTCTGGACCGCGGTTACTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 66
Qy 23 AenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 67 AACAGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 126
Qy 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db 127 GCTGTCCCTCTGGATTGAGTACCAAGCATCAAGCCCTCTTTTACCAACCTCCAT 183
RESULT 6
LOCUS CQ727777 1167 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 13711 from Patent WO02068579.
ACCESSION CQ727777
VERSION CQ727777.1 GI:42294771

KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
JOURNAL Patent: WO 02068579-A 13711 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source 1. .1167
Location/Qualifiers
/organism='Homo sapiens'
/mol_type='unassigned DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores: 2.87e-32 Length: 1167
Pred. No.: 295.00 Matches: 54
Score: 295.00
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 6 Gaps:
US-09-774-954-3 (1-61) x CQ727777 (1-1167)
Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 76 GCGGGCTCTGGACCGCGGTTACTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 135
Qy 23 AenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 136 AACAGGCGGATCACTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 195
Qy 43 AlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db 196 GCTGTCCCTCTGGATTGAGTACCAAGCATCAAGCCCTCTTTTACCAACCTCCAT 252
RESULT 7
LOCUS AJ781500 1167 bp mRNA linear PRI 07-JUL-2004
DEFINITION Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12 gene).
ACCESSION AJ781500
VERSION AJ781500.1 GI:50057069
KEYWORDS fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Martinez-Duncker,I., Mollicone,R., Candellier,J.J., Breton,C. and Oriol,R.
TITLE A new superfamily of protein-O-fucosyltransferases, alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs
JOURNAL Glycobiology 13 (12), 1C-5C (2003)
PUBMED 12986037
REFERENCE 2
AUTHORS Martinez-Duncker,I., Oriol,R. and Mollicone,R.
TITLE Phylogeny of fucosyltransferases
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1167)
AUTHORS Oriol,R.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE
FEATURES
source 1. .1167
/organism='Pan troglodytes'


```
/mol_type="mRNA"
/db_xref="taxon:9598"
1..1167
/gene="fut12"
1..1167
/gene="fut12"
/notes="adds fucose to Ser or Thr"
/notes="pofut1"
/codon_start=1
/product="protein-O-fucosyltransferase 1"
/protein_id="CAH03712.1"
/db_xref="GI:50057070"
/translations="MGAAWRLPSVFLILLPLPGMDPAGSLYLLCYPCMGGRF
GNQADFLSLAFAKLRLVAVPMLEYOHKPPFTNLHVSQYKFKLEPLQAYHRV
ISLEDFMEKLAPTHWPPEKRVAYCFEVAQSPDKTCTPMKEGNGPFPDQPHVSFN
KSLFTGISFSASRYQWSQSPFKEHPVIALPGAPQAPFVLEHRPQLQKYNWSDM
VKTGEAQIHAHLVRYPVYGIHLRIGSDWKNACMLKDGTAGSHFMAFPCQGVYSRSTAA
PLTWMLCPDKIEIQRAVKLWVRSLSDAQSYVATDESYYVPELQQLFKGKVKVSLKP
EVAQVDLYILQADHFIGNCVSSTAFVKKERDLQGRPSFFQMDRPPKLRDEF"

ORIGIN
Alignment Scores:
Pred. No.: 2,87e-32 Length: 1167
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-3 (1-61) x AJ781500 (1-1167)

Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
142 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 201
23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
202 AACACGGCGGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 261
43 AlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
262 GCTGTCCCTCTTGGATTGATTGACGAGCATCACAGCTCTCTTTCCACCAACCTCCAT 318

LOCUS
BD103189 1300 bp DNA linear PAT 27-AUG-2002
DEFINITION
O-fucosyltransferase.
ACCESSION
BD103189
VERSION
BD103189.1 GI:22648763
KEYWORDS
JP 2001527389-A/2.
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 1300)
AUTHORS
Wang,Y. and Spellman,M.W.
TITLE
O-fucosyltransferase
JOURNAL
Patent: JP 2001527389-A 2 25-DEC-2001;
GENENTECH INC
COMMENT
OS Unidentified
PN JP 2001527389-A/2
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498,26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
PC C12N15/54,C12N9/10,C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FH Key Location/Qualifiers
FT source 1..1300
FT /organism="Unidentified".
FEATURES
source
1..1300
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 3.24e-32 Length: 1300
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-3 (1-61) x BD103189 (1-1300)

Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
142 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 201
23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
202 AACACGGCGGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 261
43 AlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
262 GCTGTCCCTCTTGGATTGATTGACGAGCATCACAGCTCTCTTTCCACCAACCTCCAT 318

LOCUS
AR163451 1514 bp DNA linear PAT 17-OCT-2001
DEFINITION
Sequence 2 from patent US 6270987.
ORIGIN
Alignment Scores:
Pred. No.: 3.24e-32 Length: 1300
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-3 (1-61) x AR163452 (1-1300)

Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
142 GCGGGCTCTGGGACCGCGGTTACCTGCTCTACTGCCCTGCATGGGGCGCTTTGGG 201
23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
202 AACACGGCGGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGTACCTTG 261
43 AlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeuHis 61
262 GCTGTCCCTCTTGGATTGATTGACGAGCATCACAGCTCTCTTTCCACCAACCTCCAT 318

LOCUS
AR163452 1300 bp DNA linear PAT 17-OCT-2001
DEFINITION
Sequence 6 from patent US 6270987.
ACCESSION
AR163452
VERSION
AR163452.1 GI:16234057
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 1300)
AUTHORS
Wang,Y. and Spellman,M.W.
TITLE
O-fucosyltransferase
JOURNAL
Patent: US 6270987-A 6 07-AUG-2001;
Location/Qualifiers
FEATURES
source
1..1300
/mol_type="unknown"
/organism="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 3.24e-32 Length: 1300
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-3 (1-61) x AR163452 (1-1300)
```

ACCESSION AR163451
 VERSION AR163451.1 GI:16234056
 KEYWORDS
 SOURCE Unknown.
 ORGANISM
 Unidentified.
 Unclassified.
 1 (bases 1 to 1514)
 Wang, Y. and Spellman, M. W.
 O-fucosyltransferase
 Patent: US 6270987-A 2 07-AUG-2001;
 JOURNAL Location/Qualifiers
 FEATURES
 source
 location=1514
 /organism="unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Alignment Scores: 3.84e-32 Length: 1514
 Pred. No.: 295.00 Matches: 54
 Score: 91.53% Conservative: 0
 Percent Similarity: 91.53% Mismatches: 5
 Best Local Similarity: 91.53% Indels: 0
 Query Match: 90.49% Gaps: 0
 DB: 6
 US-09-774-954-3 (1-61) x AR163451 (1-1514)
 QY 3 AlaGlySerTipAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 Db 7 GCGGGCTCTGGGACCGCGGTTACTGCTCTACTGCGCCCTGCAATGGGCGCTTTGGG 66
 QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeu***ValArgThrLeu 42
 Db 67 AACACGGCGGATCACTTCTGGGCTCTTGGCAATTTGCAAGCTCTTAACCGTACCTTG 126
 QY 43 AlaValProProTfPileGlyTyrGlnHisGlySerProProPheThrAsnLeuHis 61
 Db 127 GCTGTCCCTCTTGGATTGAGTACCAAGCTCTTCCCTTTTCCACCACTCCAT 183
 RESULT 11
 LOCUS BD103188 1514 bp DNA linear PAT 27-AUG-2002
 DEFINITION O-fucosyltransferase.
 ACCESSION BD103188
 VERSION BD103188.1 GI:22648762
 KEYWORDS JP 2001527389-A/1.
 SOURCE unidentified
 ORGANISM
 unclassified.
 1 (bases 1 to 1514)
 Wang, Y. and Spellman, M. W.
 O-fucosyltransferase
 Patent: JP 2001527389-A 1 25-DEC-2001;
 JOURNAL GENE TECH INC
 COMMENT OS Unidentified
 PN JP 2001527389-A/1
 PD 25-DEC-2001
 PF 17-DEC-1997 JP 1998532877
 PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
 YC YANG WANG, MICHAEL W SPELLMAN
 PC C12N15/54, C12N9/10, C07K16/40
 CC Strandedness: Single;
 CC Topology: Linear;
 CC O-fucosyltransferase
 PH Key Location/Qualifiers
 FT source 1..1514
 FT Location/Qualifiers
 FT 1..1514
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 FEATURES
 source
 ORIGIN

Alignment Scores: 3.84e-32 Length: 1514
 Pred. No.: 295.00 Matches: 54
 Score: 91.53% Conservative: 0
 Percent Similarity: 91.53% Mismatches: 5
 Best Local Similarity: 91.53% Indels: 0
 Query Match: 90.49% Gaps: 0
 DB: 6
 US-09-774-954-3 (1-61) x BD103188 (1-1514)
 QY 3 AlaGlySerTipAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
 Db 7 GCGGGCTCTGGGACCGCGGTTACTGCTCTACTGCGCCCTGCAATGGGCGCTTTGGG 66
 QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeu***ValArgThrLeu 42
 Db 67 AACACGGCGGATCACTTCTGGGCTCTTGGCAATTTGCAAGCTCTTAACCGTACCTTG 126
 QY 43 AlaValProProTfPileGlyTyrGlnHisGlySerProProPheThrAsnLeuHis 61
 Db 127 GCTGTCCCTCTTGGATTGAGTACCAAGCTCTTCCCTTTTCCACCACTCCAT 183
 RESULT 12
 LOCUS BC000582 1524 bp mRNA linear PRI 29-JUN-2004
 DEFINITION Homo sapiens protein O-fucosyltransferase 1, transcript variant 2,
 mRNA (cDNA clone MGC:2482 IMAGE:3162533), complete cds.
 ACCESSION BC000582
 VERSION BC000582.2 GI:37588976
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1524)
 Strausberg, R.L., Feingold, P.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, R.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Schat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, J., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, F.J., Richards, S., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hu, Y.S., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Rogignani, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Small, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1524)
 Strausberg, R.
 Direct Submission
 Submitted (15-NOV-2000) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Oct 8, 2003 this sequence version replaced gi:12653608.
 Contact: MGC help desk
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mc@nigri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov> Series: IRAL Plate: 5 Row: 0 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27436890.

FEATURES

source

1. .1524
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:2482 IMAGE:3162533"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH MGC 19"
/lab_host="DH10B-R"
/notes="vector: pOTB7"

gene

1. .1524
/gene="POFUT1"
/notes="synonyms: KIAA0180, O-FUT, O-Fuc-T, FUT12, MGC2482, O-FucT-1"
/db_xref="LocusID:23509"
/db_xref="MIM:607491"
50. .634
/gene="POFUT1"
/codon_start=1
/product="protein O-fucosyltransferase 1, isoform 2 precursor"

CDS

precursor

/protein_id="AAH00582.1"
/db_xref="GI:12653609"
/db_xref="LocusID:23509"
/db_xref="MIM:607491"
translation="MGAAAWARPLSVSFLLLPLPGMPAGSWDPAGYLLYCPMGGRF
GNQADRLGLSLAFKLLNRLTAVPPWLEYOHKPPFTNLHVSYQYFKLPLQAYHRV
ISLEDFMEKLAPTHWPEKRVAYCFEVAAGRSDDKTCPKMKGPNPGFPFDQHFVSN
KSELFTGISFSASYREQRSQRRENHSCVTLFLFR"

ORIGIN

Alignment Scores:

Pred. No.: 3.87e-32 Length: 1524
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-3 (1-61) x BC000582 (1-1524)

Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 125 GCGGGCTCTGGAGCCGCGGTTACCTGCTACTGCTCCCTGCATGGGGCGCTTTGGG 184
Qy 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 185 AACGAGCGCATCACTTCTGGGCTCTGTCGATTTGCAAGCTGCTAAACCGTACCTTG 244
Qy 43 AlaValProTrpTrpLeuGlyTyrGlnHisLysProPheThrAsnLeuHis 61
Db 245 GCTGTCCCTCTTGGATTGAGTACGATCACAAGCCTCTCTTCCACCACTCCAT 301

RESULT 13

HSM803330

LOCUS

4560 bp mRNA linear PRI 13-MAY-2003

DEFINITION

Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).

ACCESSION

AL832023

VERSION

AL832023.1 GI:21732563

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 4560)

AUTHORS

Wambutt, R., Heubner, D., Meves, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.

TITLE

Direct Submission

JOURNAL

Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764

COMMENT

Neuberger, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp451J114) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source

1. .4560
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp451J114"
/db_xref="taxon:9606"
/clone="DKFZp451J114"
/tissue_type="human skeletal muscle"
/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
4538

polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 1.31e-31 Length: 4560
Score: 295.00 Matches: 54
Percent Similarity: 91.53% Conservative: 0
Best Local Similarity: 91.53% Mismatches: 5
Query Match: 90.49% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-3 (1-61) x HSM803330 (1-4560)

Qy 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
Db 57 GCGGGCTCTGGAGCCGCGGTTACCTGCTACTGCTCCCTGCATGGGGCGCTTTGGG 116
Qy 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
Db 117 AACGAGCGCATCACTTCTGGGCTCTGTCGATTTGCAAGCTGCTAAACCGTACCTTG 176
Qy 43 AlaValProTrpTrpLeuGlyTyrGlnHisLysProPheThrAsnLeuHis 61
Db 177 GCTGTCCCTCTTGGATTGAGTACGATCACAAGCCTCTCTTCCACCACTCCAT 233

RESULT 14

D80002

LOCUS

Homo sapiens KIAA0180 mRNA, complete cds.

DEFINITION

ACCESSION

D80002

VERSION

D80002.2 GI:20521837

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

AUTHORS

Nagase, T., Seki, N., Ishikawa, K., Tanaka, A. and Nomura, N.
Prediction of the coding sequences of unidentified human genes. V.

The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by analysis of cDNA clones from human cell line KG-1

JOURNAL
MEDLINE
PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

Chiang, P.W., Wang, S., Smithivas, P., Song, W.J., Ramamoorthy, S., Hillman, J., Puett, S., Van Keuren, M.L., Crombez, E., Kumar, A., Glover, T.W., Miller, D.B., Tsai, C.H., Blackburn, C.C., Chen, X.N., Sun, Z., Cheng, J.F., Korenberg, J.R. and Kurnit, D.M.

Identification and analysis of the human and murine putative chromatin structure regulator SUPT6H and Supt6h

Genomics 34 (3), 328-333 (1996)

96374824

8766132

3 (bases 1 to 5189)

Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.

Direct Submission

Submitted (12-DEC-1995) Osamu Ohara, Kazuo DNA Research Institute; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail:cdnaifookarusa.or.jp, Tel:+81-438-52-3913)

On May 9, 2002 this sequence version replaced gi:1136419.

Location/Qualifiers

1..5189

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="ha0256781"

/sex="male"

/cell_line="KG-1"

/cell_type="myeloblast"

/tissue_type="brain"

/note="This sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1066 was derived from pf00933 and 1067 - 5189 was derived from ha02567)."

1..5189

/gene="K1AA0180"

<1..1213

/gene="K1AA0180"

/citation={2}

/codon_start=2

/protein_id="BAJ11497.2"

/db_xref="GI:20521838"

/translation="PPRLCAAGSGRADMGAAWAPLVSFILLLLPLPGMPAGSWDPAGYLLPCPCMGPRGQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPFPTNLVSVYQYFLEPLQAYHRV

KYFLEPLQAYHRVLSLEDFMEKLAHPWPEKRVAYCFEVAQRSPDKKTCPMKEG

EHRPFGQFVHSVFNKSELTGTSFSAVREQMSQRFKHPVLALPGAPQFVLE

EHRLPKQYVMSVDSMVTGIAHILRVVYVGIHLRIGSDWKACMLKDGTFAGSHF

MASPCQVSRSTAAPLTMTMCLPDLKEIQRAVLKVRSLDAQSVVATDSESYVPEL

QQLFKGVKVSLLKPEVAQVDLYILGQADHFLGNCVSSFTAFVKRERDLQGRPSGFFG

MDRPPKLRDEF"

ORIGIN

Alignment Scores:

Pred. No.: 1 528-31 Length: 5189

Score: 295.00 Matches: 54

Percent Similarity: 91.53% Conservative: 5

Best Local Similarity: 91.53% Mismatches: 0

Query Match: 90.49% Indels: 0

DB: 9 Gaps: 0

US-09-774-954-3 (1-61) x D80002 (1-5189)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22

DB 122 GCGGGCTCTCTGGACCGCGCGGTACCTGCTCTACTGCGCGCGCTTTGGG 181

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42

DB 182 AACGAGCGCATCACTCTCTGGGCTCTCTGGCATTTGCAAGAGCTGCTAAACCGTACCTTG 241

QY 43 AlaValProTirpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61

DB 242 GCTGTCCTCTCTGGATTGAGTACCAAGCATCACAGGCTCTTTTCCACCAACCTCCAT 298

RESULT 15

AF375884

LOCUS

DEFINITION

AF375884 Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete cds.

ACCESSION

AF375884

VERSION

AF375884.1

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

REFERENCE

1 (bases 1 to 5249)

Wang, Y., Shao, Y., Shi, S., Harris, R.J., Spellman, M.W., Stanley, P.

Modification of epidermal growth factor-like repeats with O-fucose.

Molecular cloning and expression of a novel GDP-fucose protein

O-fucosyltransferase

J. Biol. Chem. 276 (43), 40338-40345 (2001)

21523965

11524432

2 (bases 1 to 5249)

Shi, S., Stanley, P., Wang, Y., Shao, Y., Harris, R.J., Spellman, M.W.

and Hattiwanger, R.S.

Direct Submission

Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of

Medicine, 1300 Morris Park Ave., New York, NY 10461, USA

Location/Qualifiers

1..5249

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="20"

/map="between FLAGL2 and KIF3B"

/tissue_type="heart"

1..5249

/gene="POFUT1"

50..1216

/gene="POFUT1"

/function="catalyzes transfer of fucose from GDP-fucose to

serine or threonine on epidermal growth factor-like

repeats"

/note="O-FucT-1, glycosyltransferase"

/codon_start=1

/product="protein o-fucosyltransferase"

/protein_id="AA09576.1"

/db_xref="GI:15825114"

/translation="NGAAAWAPLVSFLLLLPLPMPAGSWDPAGYLLPCPCMGPRF

GNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPFPTNLHVSQYFLEPLQAYHRV

ISLEDFMEKLAHPWPEKRVAYCFEVAQRSPDKKTCPMKEGNGFPFQFVHSVFN

KSELTGTSFSAVREQMSQRFKHPVLALPGAPQFVLEHRLPKQYVMSVDSM

VKTGEAQIHAHLRVVYVGIHLRIGSDWKACMLKDGTFAGSHFMASSPCQVSRSTAA

PLTMTMCLPDLKEIQRAVLKVRSLDAQSVVATDSESYVPELQQLFKGVKVSLLK

PEVAQVDLYILGQADHFLGNCVSSFTAFVKRERDLQGRPSGFFGMDRPPKLRDEF"

ORIGIN

Alignment Scores:

Pred. No.: 1 548-31 Length: 5249

Score: 295.00 Matches: 54

Percent Similarity: 91.53% Conservative: 0

Best Local Similarity: 91.53% Mismatches: 5

Query Match: 90.49% Indels: 0

DB: 9 Gaps: 0

US-09-774-954-3 (1-61) x AF375884 (1-5249)

QY 3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22

DB 125 GCGGGCTCTCTGGACCGCGCGGTACCTGCTCTACTGCGCGCGCTTTGGG 184

Qy 23 AsnGlnAlaaspHisPheLeuGlySerLeuAlaPheAlaLysLeu**ValArgThrLeu 42
 |||||
 Db 185 AACGAGCGCATCACTTCITGGGCTCTCTGGCAITTTGCAAGCTGCTAAACCGTACCTTG 244
 |||||
 Qy 43 AlaValProTrrpIleGluTyrGlnHisHisLysProPheThrAsnLeuHis 61
 |||||
 Db 245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAAGCTCCTTTTACCAACCTCCAT 301
 |||||

Search completed: October 26, 2005, 12:55:28
 Job time : 1386.21 secs

THIS PAGE BLANK (US710)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:43 ; Search time 7.64066 Seconds
(without alignments)
768.157 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMR.....LAVPPWIEYQHKKPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database ; PIR 79:.*
1: Piri:.*
2: Piri2:.*
3: Piri3:.*
4: Piri4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	35.0	474	2 T15511	hypothetical prote
2	68.5	21.0	569	2 E86957	probable cholester
3	68.5	21.0	585	2 S72824	cholesterol oxidas
4	67.5	20.7	354	2 I51266	opsin, green-sensi
5	66.5	20.4	767	2 D95101	SpoE family protei
6	66.5	20.4	767	2 E97969	cell division prot
7	63	19.3	562	2 T40967	hypothetical prote
8	62	19.0	578	2 F70736	probable chob prot
9	61.5	18.9	365	2 A36047	galactoside 2-alph
10	61	18.7	674	2 T21217	hypothetical prote
11	59.5	18.3	491	2 C70570	hypothetical prote
12	59.5	18.3	491	2 S76941	hypothetical prote
13	58.5	17.9	219	2 S72723	amidotransferase h
14	58.5	17.9	223	2 B86968	conserved hypothe
15	58.5	17.9	348	2 AE0875	D-erythrose 4-phos
16	57	17.5	363	2 T22255	hypothetical prote
17	56.5	17.3	207	2 A75475	probable acetyltra
18	56.5	17.3	225	2 T21655	hypothetical prote
19	56.5	17.3	227	2 T21662	hypothetical prote
20	56.5	17.3	355	2 A42347	opsin, green-sensi
21	56.5	17.3	467	2 AF0488	amino acid permeas
22	56.5	17.3	492	2 T41004	hypothetical prote
23	56.5	17.3	514	2 T02648	probable phospholi
24	56	17.2	214	2 C69879	hypothetical prote
25	56	17.2	328	2 A70127	phosphate ABC tran
26	55.5	17.0	158	2 A12377	lipoprotein signal
27	55.5	17.0	231	2 F83032	hypothetical prote
28	55.5	17.0	560	2 T07964	(S)-N-methylcoclau
29	55.5	17.0	1465	2 A70199	hypothetical prote

30	55	16.9	345	2 T12357	NADH2 dehydrogenas
31	55	16.9	428	2 T27763	hypothetical prote
32	54.5	16.7	346	2 T13973	NADH2 dehydrogenas
33	54.5	16.7	383	2 T14795	hypothetical prote
34	54.5	16.7	406	2 AF3342	probable aminotran
35	54.5	16.7	8563	2 T30226	polyketide synthas
36	54	16.6	344	2 T17056	NADH2 dehydrogenas
37	54	16.6	344	2 T17059	NADH2 dehydrogenas
38	54	16.6	344	2 T12335	NADH2 dehydrogenas
39	54	16.6	346	2 B90626	NADH2 dehydrogenase
40	54	16.6	346	2 T11076	NADH2 dehydrogenas
41	54	16.6	487	1 B64795	ybdS protein - Esc
42	54	16.6	487	2 G85560	probable membrane
43	54	16.6	487	2 C90710	probable membrane
44	54	16.6	501	2 AE0578	citrate carrier (i
45	53.5	16.4	333	1 PARFAS	fructose-bisphosph

ALIGNMENTS

RESULT 1

T15511
hypothetical protein C15C7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15511
R;Leimbach, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C15C7.
A;Reference number: Z18363

A;Accession: T15511
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-474 <LEI>
A;Cross-references: EMBL:U41528; NID:g1109795; PID:g1109800; PIDN:AAA83156.1; CESP:C15C7.
C;Genetics:
A;Gene: CBSP:C15C7.1
A;Introns: 31/3; 67/2; 106/2; 166/1; 202/3; 235/3; 364/3; 410/3

Query Match 35.0%; Score 114; DB 2; Length 474;
Best Local Similarity 64.7%; Pred. No. 3.2e-07;
Matches 22; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 19 GRFGNQADHFLGSLFAKLXVRTLAVPPWIEYQH 52
DB 166 GRFGNQVDQFLGVLAFALDRTLVLVLPNFIEFKH 199

RESULT 2

E86957
probable cholesterol oxidase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: E86957
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hoi
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E86957
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-569 <STO>
A;Cross-references: UNIPROT:Q9CCV1; GB:AL450380; NID:gl3092664; PIDN:CAC29897.1; GSPDB:G

Query Match 21.0%; Score 68.5; DB 2; Length 569;
Best Local Similarity 33.8%; Pred. No. 0.45;
Matches 23; Conservative 6; Mismatches 14; Indels 25; Gaps 4;

```
QY 5 SWDLACGLLYXPXMGKFGNQADHFL-----GSLAFAPKLXVRTILAVPPWIEYQ 51
DB 48 SWDLRKF-LWAPKLCYGIQRIHLNRNVMILAGVGGSLNYA-----NTLYVPP----- 97
QY 52 HKKPPPTN 59
DB 98 --EPPFN 103

RESULT 3
S72824
cholesterol oxidase (EC 1.1.3.6) choD - Mycobacterium leprae
N/Alternate names: B1620_C3_240 protein
C/Species: Mycobacterium leprae
C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: S72824
R/Smith, D.R.; Robinson, K.
submitted to the EMBL Data Library, November 1993
A/Description: Mycobacterium leprae cosmid B1620.
A/Reference number: S72584
A/Accession: S72824
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-585 <SMI>
A/Cross-references: UNIPROT:Q59530; EMBL:U00015; NID:G466931; PIDN:AAC43233.1; PID:G4669
C/Genetics:
A/Gene: choD
A/Start codon: GTG
C/Keywords: oxidoreductase

Query Match 21.0%; Score 68.5; DB 2; Length 585;
Best Local Similarity 33.8%; Pred. No. 0.47; 14; Indels 25; Gaps 4;
Matches 23; Conservative 6; Mismatches 14;

QY 5 SWDLACGLLYXPXMGKFGNQADHFL-----GSLAFAPKLXVRTILAVPPWIEYQ 51
DB 64 SWDLRKF-LWAPKLCYGIQRIHLNRNVMILAGVGGSLNYA-----NTLYVPP----- 113
QY 52 HKKPPPTN 59
DB 114 --EPPFN 119

RESULT 4
I51266
opsin, green-sensitive - Mexican tetra
C/Species: Asytanax mexicanus (Mexican tetra)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C/Accession: I51266
R/Register, E.A.; Yokoyama, B.; Yokoyama, S.
J. Mol. Evol. 39, 288-273, 1994
A/Title: Multiple origins of the green-sensitive opsin genes in fish.
A/Reference number: I51266; MUID:95018302; PMID:7952788
A/Accession: I51266
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-354 <REG>
A/Cross-references: UNIPROT:P51474; GB:S75255; NID:G807171; PIDN:AAB32221.1; PID:G807172
C/Genetics:
A/Gene: rh1IAF
A/Introns: 124/1; 180/2; 235/3; 315/3
C/Superfamily: vertebrate rhodopsin
C/Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;
F,299/Binding site: retinal (lys) (covalent) #status predicted

Query Match 20.7%; Score 67.5; DB 2; Length 354;
Best Local Similarity 29.1%; Pred. No. 0.36;
Matches 16; Conservative 6; Mismatches 28; Indels 5; Gaps 1;

QY 6 WDLA----GYLLYXPXMGKFGNQADHFLGSLAFAPKLXVRTILAVPPWIEYQHKKP 55
DB 129 WSLVLAIERIYIVVCKPMKSFKFSASHALGGIGTWTWMTCAAPPLVGSWRIYP 183
```

```
RESULT 5
D95101
SpOe family protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C/Accession: D95101
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidt,
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:2157209; PMID:11463916
A/Accession: D95101
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-767 <KUR>
A/Cross-references: UNIPROT:Q8DQ94; GB:AE005672; PIDN:AAK75005.1; PID:G14972351; GSPDB:GH
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP0878
C/Superfamily: Bacillus subtilis DNA translocase spoIIIE

Query Match 20.4%; Score 66.5; DB 2; Length 767;
Best Local Similarity 37.2%; Pred. No. 1.2;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

QY 8 LAGYLLYXPXMGKFGNQADHFLGSLAFAPKLXVRTILAVPPWIEY 50
DB 140 LIGVALYIPTAFLEFSNIGTYFIGSIL---ILVGLLVSPWSVY 179

RESULT 6
E97969
cell division protein [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C/Accession: E97969
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ek
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.; Mc
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: E97969
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-767 <KUR>
A/Cross-references: UNIPROT:Q8DQ94; GB:AE007317; PIDN:AAK99585.1; PID:G15458378; GSPDB:GH
C/Genetics:
A/Gene: ftsK
C/Superfamily: Bacillus subtilis DNA translocase spoIIIE

Query Match 20.4%; Score 66.5; DB 2; Length 767;
Best Local Similarity 37.2%; Pred. No. 1.2;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

QY 8 LAGYLLYXPXMGKFGNQADHFLGSLAFAPKLXVRTILAVPPWIEY 50
DB 140 LIGVALYIPTAFLEFSNIGTYFIGSIL---ILVGLLVSPWSVY 179

RESULT 7
T40967
hypothetical protein SPCC1450.18 - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T40967; T41000
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, October 1998
```



```
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-365 <LAR>  
C:Cross-references: UNIPROT:P19526; GB:M35531; NID:g183887; PID:g306830  
C;Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein
```

1;

Query Match 18.9%; Score 61.5; DB 2; Length 365;
Best Local Similarity 27.4%; Pred.No. 2.4; Indels 43; Gaps 5;
Matches 26; Conservative

QY 2 LAGSWDLGAYLLYPXMGREGNQADHFLGLSLAFAKLXVR-----TLA 43
| : | : |
DB 77 LSQTWTV-----YPN-GREGNGMGOVATLALAQNLGRRAFILPAMHAALAPVFRTITLP 129
| : | : |

QY 44 V-----PPWIETQH-----KPFPTNL 60
| | | | |
DB 130 VLAAPEVDSTRTPWELOLDHMWMSEYADRDRDFFKL 164
| : | : |

RESULT 10
T21217
hypothetical protein F21G4.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21217

R;Mortimore, B.
submitted to the EMBL Data Library, October 1996

A:Reference number: Z19392
A:Accession: T21217
A:Molecule type: DNA
A:Residues: 1-674 <WIL>
A:CROSS-references: UNIPROT:Q93J50; EMBL.Z81016; PIDD.CAB02665.1; GSFPDB.GN00028
A:Experimental source: clone F21G4
C:Genetics:
A:Gene: CESP:F21G4.1
A:Map position: X
A:Introns: 58/3; 149/1; 181/3; 213/3; 286/1; 327/2; 457/2; 509/3; 608/1

Query Match 18.7%; Score 61; DB 2; Length 674;
Best Local Similarity 27.4%; Pred.No. 5.4;
Matches 17; Conservative 25; Indels 6; Gaps 2

QY 4 GSWDLGAYLYXPXM-----GRFGNQADHLFGLSFAKLXVRTLVAPPWIYQQHKPPFT 58
| : | : |
DB 359 GSYLGIGYTQLPKFIETQYGRSGSMADIYSGIISVGAIIVST-ALGGWILSRNYIAPS 417
| : | : |

QY 59 NL 60
| : |

DB 418 SI 419

RESULT 11
C70570
hypothetical protein RV2604c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: C70570

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Comnor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Ho;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70570
A:Molecule type: nucleic acid sequence not shown; translation not shown
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <COL>
A:CROSS-references: UNIPROT:O06210; GB:Z95387; GB:AL123456; NID:g3261763; PIDD.N
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2604C

DB 137 FGRQVDSFEGDIGFAGLVDPVRVAVIPAPWVE 168
 ||| : | | | | | : | | | : | | | : | | | :
 RESULT 14
 B86968 conserved hypothetical protein ML0474 [imported] - Mycobacterium leprae
 C;Species: Mycobacterium leprae
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C;Accession: B86968
 R;Cole, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd,
 R.; Davies, R.M.; Devlin, K.; Dutchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 Eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq;
 A;Title: Massive gene decay in the leprosy bacillus.
 A;Reference number: A86909; MUID:21128732; PMID:11234002
 A;Accession: B86968
 A>Status: preliminary
 A:Molecule type: DNA
 A;Residues: 1-223 <STO>
 A;Cross-references: UNIPROT:Q9CCT5; GB:AL450380; NID:gl3092704; PIDN:CAC29982.1; GSPDB:GN00176
 C;Genetics:
 A;Gene: ML0474
 C;Superfamily: conserved hypothetical protein H11648

Query Match 17.9%; Score 58.5; DB 2; Length 223;
 Best Local Similarity 43.8%; Pred. No. 3.4;
 Matches 14; Conservative 4; Mismatches 11; Indels 3; Gaps 2;

QY 21 FGNQADHFLGSLAFAPLX--VRTLAV-PPWIE 49
 ||| : | | | | | : | | | : | | | : | | | :
 DB 141 FGRQVDSFEGDIGFAGLVDPVRVAVIPAPWVE 172
 ||| : | | | | | : | | | : | | | : | | | :

RESULT 15
 AE0875 D-erythrose 4-phosphate dehydrogenase [imported] - Salmonella enterica subsp. enterica sero
 C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: This species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AE0875
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
 A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AE0875
 A>Status: preliminary
 A:Molecule type: DNA
 A;Residues: 1-348 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD02901.1; PID:g16504154; GSPDB:GN00176
 C;Genetics:
 A;Gene: STY3228
 C;Superfamily: glyceraldehyde-3-phosphate dehydrogenase

Query Match 17.9%; Score 58.5; DB 2; Length 348;
 Best Local Similarity 35.8%; Pred. No. 5.6;
 Matches 19; Conservative 4; Mismatches 25; Indels 5; Gaps 1;

QY 2 LAGSWDLAgyLLYPXMGFGNOADH-----FLGSLAFAPLXVETLAVPPWIE 49
 ||| : | | | | | : | | | : | | | : | | | :
 DB 38 LADAAGMAHLKYDTSHGFAWEVHERERQLFVGDDVIRILHRTLDLPWE 90
 ||| : | | | | | : | | | : | | | : | | | :

Search completed: October 25, 2005, 15:36:43
 Job time : 9.64066 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 38.8296 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLASGDLAGLYLXPXGR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	98.2	61	2 AAW80572	Aaw80572 N-termina
2	295	90.5	61	2 AAW80574	Aaw80574 N-termina
3	295	90.5	365	2 AAW80571	Aaw80571 Human hea
4	295	90.5	388	8 ABM82223	Abm82223 Tumour-as
5	295	90.5	397	2 AAW80573	Aaw80573 Human O-f
6	232.5	71.3	417	4 AAW78377	Aaw78377 Human pro
7	220.5	67.6	417	4 AAW79361	Aaw79361 Human pro
8	204	62.6	343	2 AAW80577	Aaw80577 Partial h
9	177	54.3	402	4 ABB63790	Abb63790 Drosophil
10	81	24.8	28	2 AAW80575	Aaw80575 N-termina
11	76	23.3	150	3 AAY87325	Aay87325 Human sig
12	69.5	21.3	198	6 ABU33866	Abu33866 Protein e
13	66.5	20.4	737	8 ADK48560	Adk48560 Streptoco
14	66.5	20.4	741	6 ABU01274	Abu01274 S. pneumo
15	66.5	20.4	763	8 ADR94975	Adr94975 Novel S.
16	66.5	20.4	767	4 AAU37750	Aau37750 Streptoco
17	66.5	20.4	767	6 AAU38005	Aau38005 Streptoco
18	66.5	20.4	767	6 ABU45987	Abu45987 Protein e
19	66.5	20.4	768	3 AAY81596	Aay81596 Streptoco
20	66.5	20.4	783	2 AAW97416	Aaw97416 A membran
21	61.5	18.9	365	2 AAR13751	Aar13751 GDP-Fuc:b
22	61.5	18.9	365	2 AAR45936	Aar45936 A glycosy
23	61.5	18.9	365	2 AAR70421	Aar70421 2-Alpha-f
24	61.5	18.9	365	2 AAR80154	Aar80154 GDP-L-fuc
25	61.5	18.9	365	2 AAR70422	Aar70422 2-Alpha-f

26	61.5	18.9	365	2 AAR90572	Aar90572 Human H-t
27	61.5	18.9	365	2 AAW23805	Aaw23805 Human alp
28	61.5	18.9	365	2 AAW13640	Aaw13640 Human alp
29	61.5	18.9	365	3 AAY97279	Aay97279 Human H-t
30	61.5	18.9	365	5 AAE29230	Aae29230 Human alp
31	60.5	18.6	580	8 ADN25627	Adn25627 Bacterial
32	59.5	18.3	198	6 ABU36823	Abu36823 Protein e
33	59.5	18.3	198	6 ABU34438	Abu34438 Protein e
34	59.5	18.3	340	2 AAR63066	Aar63066 Acetyl-es
35	59	18.1	366	6 ABM69163	Abm69163 Photorhab
36	58.5	17.9	223	6 ABU35781	Abu35781 Protein e
37	58.5	17.9	365	2 AAW53102	Aaw53102 Pig H tra
38	58.5	17.9	365	2 AAW30630	Aaw30630 Swine alp
39	58.5	17.9	365	2 AAW97356	Aaw97356 Swine alp
40	58.5	17.9	365	3 AAY79302	Aay79302 Pig alpha
41	58.5	17.9	365	5 AAB47995	Aab47995 Swine alp
42	57	17.5	322	8 ADR88379	Adr88379 Caenorhab
43	57	17.5	363	8 ADN23009	Adn23009 Bacterial
44	56.5	17.3	514	5 ABB91896	Abb91896 Herbicida
45	56	17.2	245	5 ABP42996	Abp42996 Human ova

ALIGNMENTS

RESULT 1
AAW80572
ID AAW80572 standard; peptide; 61 AA.
XX
AC AAW80572;
XX
DT 16-DEC-1998 (first entry)
XX
DE N-terminal amino acid sequence of CHO O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 15 /note= "unknown"
FT Misc-difference 17 /note= "unknown"
FT Misc-difference 38 /note= "unknown"
FT
XX WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX WPI; 1998-437477/37.
XX
PT Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 5; Page 69; 90pp; English.
XX
CC This represents the N-terminal sequence of CHO O-fucosyltransferase. The
CC enzyme can glycosylate an epidermal growth factor (EGF) domain of a
CC polypeptide with an activated O-fucose residue. Inhibitors of O-
CC fucosyltransferase, e.g. mutants with increased affinity for the EFG

CC domains, are used in diagnosis and treatment of conditions associated
 CC with overexpression of O-fucosyltransferase, to promote survival of
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
 CC used to detect gene amplification and expression. The expression can also
 CC be determined at the protein level using antibodies specific for O-
 CC fucosyltransferase

XX SQ Sequence 61 AA;

Query Match 98.2%; Score 320; DB 2; Length 61;
 Best Local Similarity 100.0%; Pred. No. 9.3e-38;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAGSWDLAGYLLYXPXMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNL 60
 |||||
 Db 1 RLAGSWDLAGYLLYXPXMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNL 60

QY: 61 H 61

Db 61 H 61

RESULT 2

AAW80574
 ID AAW80574 standard; peptide; 61 AA.

XX AC AAW80574;

XX DT 16-DEC-1998 (first entry)

XX DE N-terminal amino acid sequence of human heart O-fucosyltransferase.

XX KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
 XX KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.

XX OS Homo sapiens.

XX PN WO9833924-A1.

XX PD 06-AUG-1998.

XX PF 17-DEC-1997; 97WO-US023401.

XX PR 31-JAN-1997; 97US-00792498.

XX PR 26-NOV-1997; 97US-00978741.

XX PA (GETH) GENENTECH INC.

XX PI Wang Y, Spellman MW;

XX DR WPI; 1998-437477/37.

XX PT Human O-fucosyltransferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.

XX PS Claim 4; Page 69; 90pp; English.

XX CC This represents a the N-terminal sequence of the human heart O-
 CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
 CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
 CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
 CC domains, are used in diagnosis and treatment of conditions associated
 CC with overexpression of O-fucosyltransferase, to promote survival of
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
 CC used to detect gene amplification and expression. The expression can also
 CC be determined at the protein level using antibodies specific for O-
 CC fucosyltransferase

XX SQ Sequence 61 AA;

Query Match 90.5%; Score 295; DB 2; Length 61;

Best Local Similarity 91.5%; Pred. No. 3.5e-34;

XX ID ABM82223 standard; protein; 388 AA.

XX

Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNLH 61
 |||||

Db 3 AGSWDPAGYLLYCPQMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNLH 61
 |||||

RESULT 3

AAW80571
 ID AAW80571 standard; protein; 365 AA.

XX AC AAW80571;

XX DT 16-DEC-1998 (first entry)

XX DE Human heart O-fucosyltransferase.

XX KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
 XX KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.

XX OS Homo sapiens.

XX PI Wang Y, Spellman MW;

XX DR WPI; 1998-437477/37.

XX DR N-PSDB; AAW65632.

XX PT Human O-fucosyltransferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.

XX PS Example 1; Fig 12A; 90pp; English.

XX CC This represents a human heart O-fucosyltransferase that can glycosylate
 CC an epidermal growth factor (EGF) domain of a polypeptide with an
 CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
 CC mutants with increased affinity for the EGF domains, are used in
 CC diagnosis and treatment of conditions associated with overexpression of O-
 CC fucosyltransferase, to promote survival of sensory (retinal) neurons.
 CC Probes based on EGF domain polypeptide are used to detect gene
 CC amplification and expression. The expression can also be determined at
 CC the protein level using antibodies specific for O-fucosyltransferase

XX SQ Sequence 365 AA;

Query Match 90.5%; Score 295; DB 2; Length 365;

Best Local Similarity 91.5%; Pred. No. 2.9e-33;

Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNLH 61
 |||||

Db 3 AGSWDPAGYLLYCPQMGREGNQADHFLGSLAFAKLVRTLAVPPWIEYQHHKPPFTNLH 61
 |||||

RESULT 4

ABM82223
 ID ABM82223 standard; protein; 388 AA.

XX

AC ABM82223;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key
FT Peptide
FT Location/Qualifiers
FT 27..32
FT /note= "polyhistidine tag"
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.
XX
XX 02-OCT-2002; 2002US-0414971P.
XX
XX (GETH) GENENTECH INC.
XX
XX Wu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX
XX N-PSDB; ACN40746.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 5726; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT polypeptide of the invention
XX
SQ Sequence 388 AA;

Query Match 90.5%; Score 295; DB 8; Length 388;
Best Local Similarity 91.5%; Pred. No. 3.1e-33;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGLYLLYXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 61
DB 26 AGSWDPAGLYLLYPCPCMGRCFGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 84

RESULT 5
AAW80573
ID AAW80573 standard; protein; 397 AA.

XX AAW80573;
XX
XX 16-DEC-1998 (first entry)
XX
XX Human O-fucosyltransferase sequence expressed by a plasmid insertion.
XX
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key
XX Location/Qualifiers
XX 27..32
XX /note= "polyhistidine tag"
XX
XX WO9833924-A1.
XX
XX 06-AUG-1998.
XX
XX 17-DEC-1997; 97WO-US023401.
XX
XX 31-JAN-1997; 97US-00792498.
XX
XX 26-NOV-1997; 97US-00978741.
XX
XX (GETH) GENENTECH INC.
XX
XX Wang Y, Spellman MW;
XX
XX WPI; 1998-437477/37.
XX
XX N-PSDB; AAV65633.
XX
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
XX domains - useful for diagnosis and treatment of diseases involving
XX overexpression of the enzyme.
XX
XX Example; Fig 13B; 90pp; English.
XX
XX This represents the amino acid sequence of the human heart O-
XX fucosyltransferase expressed by a plasmid insertion. The human O-
XX fucosyltransferase can glycosylate an epidermal growth factor (EGF)
XX domain of a polypeptide with an activated O-fucose residue. Inhibitors of
XX O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
XX domains, are used in diagnosis and treatment of conditions associated
XX with overexpression of O-fucosyltransferase, to promote survival of
XX sensory (retinal) neurons. Probes based on EGF domain polypeptide are
XX used to detect gene amplification and expression. The expression can also
XX be determined at the protein level using antibodies specific for O-
XX fucosyltransferase
XX
SQ Sequence 397 AA;

Query Match 90.5%; Score 295; DB 2; Length 397;
Best Local Similarity 91.5%; Pred. No. 3.2e-33;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGLYLLYXPMGRFGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 61
DB 35 AGSWDPAGLYLLYPCPCMGRCFGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 93

RESULT 6
AAW78377
ID AAW78377 standard; protein; 417 AA.
XX
XX AAW78377;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1039.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

PN WO9833924-A1.
 XX
 PD 06-AUG-1998.
 XX
 XX 17-DEC-1997; 97WO-US023401.
 XX
 XX 31-JAN-1997; 97US-00792498.
 PR 26-NOV-1997; 97US-00978741.
 XX
 XX (GETH) GENENTECH INC.
 XX
 XX Wang Y, Spellman MW;
 XX WPI; 1998-437477/37.
 DR N-PSDB; AAV65634.
 XX
 XX Human O-fucosyl:transferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.
 XX
 XX Example 1; Page 62-68; 90pp; English.
 XX
 XX This represents a published partial human sequence of unknown function
 CC from a myeloblast cell line. The invention provides a human heart O-
 CC fucosyltransferase enzyme that can glycosylate an epidermal growth factor
 CC (EGF) domain of a polypeptide with an activated O-fucose residue.
 CC Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
 CC for the EGF domains, are used in diagnosis and treatment of conditions
 CC associated with overexpression of O-fucosyltransferase, to promote
 CC survival of sensory (retinal) neurons. Probes based on EGF domain
 CC polypeptide are used to detect gene amplification and expression. The
 CC expression can also be determined at the protein level using antibodies
 CC specific for O-fucosyltransferase
 XX
 XX Sequence 343 AA;
 SQ
 Query Match 62.6%; Score 204; DB 2; Length 343;
 Best Local Similarity 94.9%; Pred. No. 2.8e-20;
 Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 23 NQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 61
 DB 1 NQADHFLGSLAFKLVRTLAVPPWIEYQHKKPFTNLH 39
 XX
 RESULT 9
 ABB63790
 ID ABB63790 standard; protein; 402 AA.
 AC ABB63790;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 18162.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX

DR WPI; 2001-656860/75.
 XX N-PSDB; ABL07893.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 402 AA;
 SQ
 Query Match 54.3%; Score 177; DB 4; Length 402;
 Best Local Similarity 73.3%; Pred. No. 2.5e-16;
 Matches 33; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 XX
 QY 7 DLAGVLLYXPXMGRCNQADHFLGSLAFKLVRTLAVPPWIEYQ 51
 DB 28 DPNGLTYPCMGRCNQADHFLGSLAFKLVRTLAVPPWIEYR 72
 XX
 RESULT 10
 AAW80575
 ID AAW80575 standard; protein; 28 AA.
 XX
 XX AAW80575;
 AC
 XX 16-DEC-1998 (first entry)
 DT
 XX N-terminal sequence of expressed human O-fucosyltransferase.
 DE
 XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
 KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
 XX
 OS Homo sapiens.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 23 /note= "unknown"
 FT Misc-difference 25 /note= "unknown"
 FT
 XX WO9833924-A1.
 XX
 XX 06-AUG-1998.
 PD
 XX 17-DEC-1997; 97WO-US023401.
 PF
 XX 31-JAN-1997; 97US-00792498.
 PR 26-NOV-1997; 97US-00978741.
 XX
 XX (GETH) GENENTECH INC.
 PA
 XX Wang Y, Spellman MW;
 XX WPI; 1998-437477/37.
 DR
 XX Human O-fucosyl:transferase able to glycosylate epidermal growth factor
 PT domains - useful for diagnosis and treatment of diseases involving
 PT overexpression of the enzyme.
 XX
 XX Example 1; Page 39; 90pp; English.
 PS

XX This represents the N-terminal amino acid sequence of the human heart O-
 CC fucosyltransferase expressed by a plasmid insertion. The human O-
 CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
 CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
 CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
 CC domains, are used in diagnosis and treatment of conditions associated
 CC with overexpression of O-fucosyltransferase, to promote survival of
 CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
 CC used to detect gene amplification and expression. The expression can also
 CC be determined at the protein level using antibodies specific for O-
 CC fucosyltransferase
 XX
 SQ Sequence 28 AA;
 Query Match 24.8%; Score 81; DB 2; Length 28;
 Best Local Similarity 94.4%; Pred. No. 0.00057;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 AGSWDLAGVLLYXPXMR 20
 DB 11 AGSWDPAGVLLYXPXMR 28
 ||||| ||||| ||||| |||||
 11-MAY-2000 (first entry)
 Human signal peptide containing protein HSPP-102 SEQ ID NO:102.
 XX
 DE Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
 KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;
 KW antiaesthetic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;
 KW muscular dystrophy.
 XX
 OS Homo sapiens.
 XX
 XX WO200000610-A2.
 XX
 XX 06-JAN-2000.
 XX
 XX 25-JUN-1999; 99WO-US014484.
 XX
 XX 26-JUN-1998; 98US-0090762P.
 XX
 XX 31-JUL-1998; 98US-0094983P.
 XX
 XX 01-OCT-1998; 98US-0102686P.
 XX
 XX 11-DEC-1998; 98US-0112129P.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
 XX Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
 XX Bandman O;
 XX
 XX WPI: 2000-160673/14.
 XX
 XX N-PSDB; AAZ96210.
 XX
 XX New human signal peptide-containing proteins useful in treatment,
 XX prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
 XX disease.
 XX
 XX Claim 1; Page 225-226; 327pp; English.
 XX
 XX AA298109 to AA298242 encode AA298224 to AA298357 which represent the

CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
 CC anticancer, anti-inflammatory, antimicrobial, neurotropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities and can be
 CC used in gene therapy. HSPPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSPP. Antagonists of
 CC HSPP are used to treat or prevent disorders associated with increased
 CC activity or function of HSPP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
 CC nucleic acids can be used for the recombinant production of HSPP, for
 CC detecting HSPP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
 CC or ribozyme therapeutics, for detecting related sequences or genetic
 CC variations, and for chromosomal mapping. HSPP are also used to raise
 CC specific antibodies (Ab) and to screen for agonists and antagonists
 CC (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
 CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
 CC competitive drug screens, and for purification of HSPP from natural
 CC sources
 XX
 SQ Sequence 150 AA;
 Query Match 23.3%; Score 76; DB 3; Length 150;
 Best Local Similarity 77.8%; Pred. No. 0.022;
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 3 AGSWDLAGVLLYXPXMR 20
 DB 26 AGSWDPAGVLLYPCQMGK 43
 ||||| ||||| ||||| |||||
 RESULT 12
 ABU33866
 ID ABU33866 standard; protein; 198 AA.
 XX
 XX ABU33866;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #19393.
 XX
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Mycobacterium avium.
 OS
 XX WO20027183-A2.
 XX
 XX 03-OCT-2002.
 XX
 XX 21-MAR-2002; 2002WO-US009107.
 XX
 XX 21-MAR-2001; 2001US-00815242.
 XX
 XX 06-SEP-2001; 2001US-00948993.
 XX
 XX 25-OCT-2001; 2001US-0342923P.
 XX
 XX 08-FEB-2002; 2002US-00072851.
 XX
 XX 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 XX WPI: 2003-029926/02.
 XX
 XX N-PSDB; ACA37736.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 61790; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 198 AA;

Query Match 21.3%; Score 69.5; DB 6; Length 198;

Best Local Similarity 50.0%; Pred. No. 0.25; Mismatches 3; Gaps 1;

Matches 16; Conservative 2;

QY 21 FGNQADHFLGSLAFKLVKRTLAV---PPWIE 49

DB 116 FGRQVDSFEGDIAFAGLDVPVRAVPIRAPWVE 147

RESULT 13

ADK48560

ID ADK48560 standard; protein; 737 AA.

AC ADK48560;

XX 20-MAY-2004 (first entry)

XX Streptococcus pneumoniae protein, Seq ID No 5075.

DE Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX US6699703-B1.

XX 02-MAR-2004.

XX 26-MAY-2000; 2000US-00583110.

XX 02-JUL-1997; 97US-0051553P.

XX 12-MAY-1998; 98US-0085131P.

XX 30-JUN-1998; 98US-00107433.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;

DR WPI; 2004-212399/20.

DR N-PSDB; ADK45899.

XX New nucleic acid molecules and polypeptides useful for diagnosing,

PT preventing and treating pathological conditions resulting from bacterial

PT infection, e.g. Streptococcus pneumoniae infection, and in drug

PT screening.

XX Disclosure; SEQ ID NO 5075; 301pp; English.

XX The invention relates to isolated Streptococcus pneumoniae nucleic acids

CC and polypeptides. The nucleic acids and proteins are useful for

CC diagnosing, preventing and treating pathological conditions resulting

CC from bacterial infection, such as *S. pneumoniae* infection. These may also

CC be used for drug screening procedures. The present sequence represents a

CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence

CC data for this patent did not appear in the printed specification but was

CC obtained in electronic format directly from USPTO at

CC seqdata.uspto.gov/sequence.html.

XX SQ Sequence 737 AA;

Query Match 20.4%; Score 66.5; DB 8; Length 737;

Best Local Similarity 37.2%; Pred. No. 3.2;

Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

QY 8 LAGLYLXXPMGFGNQADHFLGSLAFKLVKRTLAVPPWIEY 50

DB 140 LIGVALYIPTAFILFSGNIGTYFISIL---ILVGSLLVSPWSVY 179

RESULT 14

ABU01274

ID ABU01274 standard; protein; 741 AA.

XX AC ABU01274;

XX 23-OCT-2003 (revised)

DT 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain protein from coding region #849.

XX Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;

KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;

KW gene therapy; vaccine.

XX Streptococcus pneumoniae; type 4 strain.

XX WO200277021-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-IB002163.

XX 27-MAR-2001; 2001GB-00007658.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelin H, Fraser C;

XX WPI; 2003-040579/03.

DR N-PSDB; ABX06561.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,

PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or

PT ear infection.

XX Claim 1; SEQ ID NO 1698; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%

CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pt_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ

Query Match 20.4%; Score 66.5; DB 6; Length 741;
Best Local Similarity 37.2%; Pred. No. 3.2;
Matches 16; Conservative 5; Mismatches 3; Gaps 1;
QY 8 LAGVLLYXPMGRFGNQADHFLGSLAFKLVKTLAVPPWIEY 50
DB 114 LIGVALYIETAPLFNSICTYFIGSIL---ILVGSLLVSPWSVY 153

RESULT 15
ADR94975
ID ADR94975 standard; protein; 763 AA.
XX
AC ADR94975;
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 3610.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
KW bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2004-697205/68.
DR N-PSDB; ADR92372.
XX
XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating

PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 3610; 151pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequences.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ

Query Match 20.4%; Score 66.5; DB 8; Length 763;
Best Local Similarity 37.2%; Pred. No. 3.4;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;
QY 8 LAGVLLYXPMGRFGNQADHFLGSLAFKLVKTLAVPPWIEY 50
DB 166 LIGVALYIETAPLFNSICTYFIGSIL---ILVGSLLVSPWSVY 205

Search completed: October 25, 2005, 15:30:40
Job time : 41.8296 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 10.0205 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLGSWDLAGLYLXXPMGR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	320	98.2	61	3	US-08-978-741-3
2	320	98.2	61	3	US-09-333-729A-5
3	295	90.5	61	3	US-08-978-741-9
4	295	90.5	61	3	US-09-333-729A-4
5	295	90.5	365	3	US-08-978-741-2
6	295	90.5	365	3	US-09-333-729A-3
7	295	90.5	397	3	US-08-978-741-6
8	295	90.5	397	3	US-09-333-729A-7
9	204	62.6	343	3	US-08-978-741-17
10	204	62.6	343	3	US-09-333-729A-13
11	177	54.3	399	4	US-09-270-767-45921
12	114	35.0	474	3	US-08-978-741-8
13	114	35.0	474	3	US-09-333-729A-12
14	81	24.8	28	3	US-08-978-741-14
15	81	24.8	28	3	US-09-333-729A-14
16	66.5	20.4	737	4	US-09-583-110-5075
17	66.5	20.4	783	2	US-09-107-433-3610
18	66.5	20.4	783	2	US-08-922-837-2
19	66.5	20.4	783	3	US-09-351-550-2
20	61.5	18.9	333	5	PCT-US91-00899-11
21	61.5	18.9	365	1	US-07-914-281-6
22	61.5	18.9	365	1	US-08-393-246-6
23	61.5	18.9	365	1	US-08-273-411-1
24	61.5	18.9	365	1	US-08-525-058A-6
25	61.5	18.9	365	1	US-08-395-800A-6
26	61.5	18.9	365	1	US-08-395-800A-10
27	61.5	18.9	365	2	US-08-696-731-6

28	61.5	18.9	365	3	US-09-042-531-6	Sequence 6, Appli
29	61.5	18.9	365	3	US-09-254-077A-11	Sequence 11, Appl
30	61.5	18.9	365	5	PCT-US91-00899-12	Sequence 12, Appl
31	61.5	18.9	399	4	US-09-949-016-11353	Sequence 11353, A
32	59.5	18.3	340	2	US-08-578-592-5	Sequence 5, Appli
33	59.5	18.3	340	3	US-09-185-111-5	Sequence 5, Appli
34	58.5	17.9	365	3	US-09-151-592-2	Sequence 2, Appli
35	58.5	17.9	365	3	US-09-254-077A-10	Sequence 10, Appl
36	58.5	17.9	365	4	US-09-443-766-13	Sequence 13, Appl
37	56	17.2	7257	3	US-09-335-409-5	Sequence 5, Appli
38	56	17.2	7257	3	US-09-568-102-5	Sequence 5, Appli
39	56	17.2	7257	3	US-09-567-969-5	Sequence 5, Appli
40	56	17.2	7257	3	US-09-568-480-5	Sequence 5, Appli
41	56	17.2	7257	3	US-09-568-486-5	Sequence 5, Appli
42	56	17.2	7257	3	US-09-568-472-5	Sequence 5, Appli
43	56	17.2	7257	3	US-09-567-899-5	Sequence 5, Appli
44	55.5	17.0	340	3	US-09-254-077A-6	Sequence 6, Appli
45	55.5	17.0	340	3	US-09-254-077A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-978-741-3
; Sequence 3, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-3

Query Match 98.2%; Score 320; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLGSWDLAGLYLXXPMGRGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNL 60
Db 1 RLGSWDLAGLYLXXPMGRGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNL 60
Qy 61 H 61

```
;
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-9

Query Match
Best Local Similarity 90.5%; Score 295; DB 3; Length 61;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGFRGNQADHFLGSLAFAPAKLVRTLAVPPWIEYQHHKPPFTNLH 61
Db 3 AGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAPAKLVRTLAVPPWIEYQHHKPPFTNLH 61

RESULT 4
US-09-333-729A-4
; Sequence 4, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 4
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-333-729A-4

Query Match
Best Local Similarity 90.5%; Score 295; DB 3; Length 61;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGFRGNQADHFLGSLAFAPAKLVRTLAVPPWIEYQHHKPPFTNLH 61
Db 3 AGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAPAKLVRTLAVPPWIEYQHHKPPFTNLH 61

RESULT 5
US-08-978-741-2
; Sequence 2, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-Nov. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: 08/792498
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
```

```
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
US-08-978-741-2

Query Match          90.5%; Score 295; DB 3; Length 365;
Best Local Similarity 91.5%; Pred. No. 1.8e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 3 AGSWDPAGYLLYCPCMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61

RESULT 6
US-09-333-729A-3
; Sequence 3, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 3
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo Sapien
;
US-09-333-729A-3

Query Match          90.5%; Score 295; DB 3; Length 365;
Best Local Similarity 91.5%; Pred. No. 1.8e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 3 AGSWDPAGYLLYCPCMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61

RESULT 7
US-08-978-741-6
; Sequence 6, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
US-08-978-741-6

Query Match          90.5%; Score 295; DB 3; Length 397;
Best Local Similarity 91.5%; Pred. No. 2e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 35 AGSWDPAGYLLYCPCMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 93

RESULT 8
US-09-333-729A-7
; Sequence 7, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert encoded protein.
;
US-09-333-729A-7

Query Match          90.5%; Score 295; DB 3; Length 397;
Best Local Similarity 91.5%; Pred. No. 2e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
DB 35 AGSWDPAGYLLYCPCMGRCGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 93

RESULT 9
US-08-978-741-17
; Sequence 17, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
```

```

; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-17
; Query Match 62.6%; Score 204; DB 3; Length 343;
; Best Local Similarity 94.9%; Pred. No. 2.8e-19;
; Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 23 NOADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNLH 61
Db 1 NOADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNLH 39
RESULT 10
US-09-333-729A-13
; Sequence 13, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang,Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 13
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-333-729A-13
; Query Match 62.6%; Score 204; DB 3; Length 343;
; Best Local Similarity 94.9%; Pred. No. 2.8e-19;
; Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 23 NOADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNLH 61
Db 1 NOADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNLH 39
RESULT 11
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.

```

Search completed: October 25, 2005, 15:42:48
Job time : 11.0205 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 35.8234 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-3

Perfect score: 326

Sequence: 1 RLASGNDLAGLLYXPXMR.....LAVPPWIEYQHKKPFTNLH 61

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	98.2	61	11	US-09-774-954-3
2	295	90.5	61	11	Sequence 3, Appli
3	295	90.5	365	11	Sequence 9, Appli
4	295	90.5	388	14	Sequence 2, Appli
5	295	90.5	397	11	Sequence 169, App
6	204	62.6	343	11	Sequence 6, Appli
7	177	54.3	402	20	Sequence 17, Appl
8	114	35.0	474	11	Sequence 18162, A
9	81	24.8	28	11	Sequence 8, Appli
10	76	23.3	150	18	Sequence 14, Appli
11	72.5	22.2	601	14	Sequence 102, App
					Sequence 12551, A

12 71.5 21.9 189 16 US-10-425-115-201329
13 69.5 21.3 198 15 US-10-282-122A-61790
14 68.5 21.0 152 15 US-10-424-599-266118
15 67.5 20.7 538 16 US-10-437-963-113919
16 66.5 20.4 741 17 US-10-472-928-1698
17 66.5 20.4 763 18 US-10-617-320-3610
18 66.5 20.4 767 9 US-09-815-242-13343
19 66.5 20.4 767 9 US-09-815-242-13598
20 66.5 20.4 767 10 US-09-769-787-96
21 66.5 20.4 767 15 US-10-282-122A-73911
22 66.5 20.4 783 9 US-09-775-978-2
23 61.5 18.9 365 9 US-09-254-077A-11
24 61.5 18.9 365 9 US-09-863-475A-6
25 61.5 18.9 365 14 US-10-105-963-10
26 61 18.7 75 16 US-10-437-963-138684
27 61 18.7 117 16 US-10-425-115-235644
28 60.5 18.6 580 15 US-10-369-493-8280
29 59.5 18.3 198 15 US-10-282-122A-64747
30 59.5 18.3 241 15 US-10-424-599-273532
31 59.5 18.3 126 15 US-10-424-599-219171
32 59 18.1 335 14 US-10-156-761-11617
33 59 18.1 223 15 US-10-282-122A-63705
34 58.5 17.9 365 9 US-09-051-034A-4
35 58.5 17.9 365 9 US-09-254-077A-10
36 58.5 17.9 365 9 US-09-844-268-13
37 58.5 17.9 365 9 US-09-844-705-13
38 58.5 17.9 110 16 US-10-425-115-327031
39 58 17.8 587 15 US-10-424-599-179849
40 57.5 17.6 322 16 US-10-664-421-10
41 57 17.5 322 16 US-10-377-268-18
42 57 17.5 322 18 US-10-941-635-10
43 57 17.5 363 15 US-10-369-493-5662
44 57 17.5 376 17 US-10-732-923-19411
45 57 17.5

ALIGNMENTS

RESULT 1

US-09-774-954-3
; Sequence 3, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:

Sequence 201329, A
Sequence 61790, A
Sequence 266118, A
Sequence 113919, A
Sequence 1698, Ap
Sequence 3610, Ap
Sequence 13343, A
Sequence 13598, A
Sequence 96, Appli
Sequence 73911, A
Sequence 2, Appli
Sequence 11, Appli
Sequence 6, Appli
Sequence 10, Appli
Sequence 198684, A
Sequence 235644, A
Sequence 8280, Ap
Sequence 62362, A
Sequence 64747, A
Sequence 273532, A
Sequence 219171, A
Sequence 11617, A
Sequence 63705, A
Sequence 4, Appli
Sequence 10, Appli
Sequence 13, Appli
Sequence 13, Appli
Sequence 327031, A
Sequence 179849, A
Sequence 10, Appli
Sequence 18, Appli
Sequence 10, Appli
Sequence 5662, Ap
Sequence 19411, A

;; CURRENT APPLICATION NUMBER: US/10/301,822
;; CURRENT FILING DATE: 2002-11-21
;; PRIOR APPLICATION NUMBER: US 60/339,971
;; PRIOR FILING DATE: 2001-12-10
;; PRIOR APPLICATION NUMBER: US 60/361,978
;; PRIOR FILING DATE: 2002-03-05
;; PRIOR APPLICATION NUMBER: US 60/381,988
;; PRIOR FILING DATE: 2002-05-20
;; NUMBER OF SEQ ID NOS: 228
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 169
;; LENGTH: 388
;; TYPE: PRT
;; ORGANISM: Homo Sapiens
US-10-301-822-169

Query Match 90.5%; Score 295; DB 14; Length 388;
Best Local Similarity 91.5%; Pred. No. 6.7e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 AGSWDLAGYLLYXPXMGREFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
Db 26 AGSWDPAGYLLYPCPCMGREFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 84

RESULT 5
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:

;; APPLICANT: Yang Wang, Michael W. Spellman
;; TITLE OF INVENTION: O-Fucosyltransferase
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 1 DNA Way
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WinPatIn (Genentech)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/774,954
;; FILING DATE: 30-Jan-2001
;; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/978,741
;; FILING DATE: 26-NOV-1997
;; APPLICATION NUMBER: 08/792,498
;; FILING DATE: 31-JAN-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Svoboda, Craig G.
;; REGISTRATION NUMBER: 39,044
;; REFERENCE/DOCKET NUMBER: P1041P1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/225-1489
;; TELEFAX: 650/952-9881

;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 397 amino acids
;; TYPE: Amino Acid
;; TOPOLOGY: Linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Query Match 90.5%; Score 295; DB 11; Length 397;
Best Local Similarity 91.5%; Pred. No. 6.9e-31;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 AGSWDLAGYLLYXPXMGREFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
Db 35 AGSWDPAGYLLYPCPCMGREFGNQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 93

RESULT 6

US-09-774-954-17
; Sequence 17, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:

;; APPLICANT: Yang Wang, Michael W. Spellman
;; TITLE OF INVENTION: O-Fucosyltransferase
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 1 DNA Way
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WinPatIn (Genentech)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/774,954
;; FILING DATE: 30-Jan-2001
;; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/978,741
;; FILING DATE: 26-NOV-1997
;; APPLICATION NUMBER: 08/792,498
;; FILING DATE: 31-JAN-1997

;; ATTORNEY/AGENT INFORMATION:
;; NAME: Svoboda, Craig G.
;; REGISTRATION NUMBER: 39,044
;; REFERENCE/DOCKET NUMBER: P1041P1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/225-1489
;; TELEFAX: 650/952-9881

;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 343 amino acids
;; TYPE: Amino Acid
;; TOPOLOGY: Linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-09-774-954-17

Query Match 62.6%; Score 204; DB 11; Length 343;
Best Local Similarity 94.9%; Pred. No. 9.5e-19;
Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 23 NQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 61
Db 1 NQADHFLGSLAFKLVRTLAVPPWIEYQHHKPPFTNLH 39

RESULT 7

US-11-097-143-18162
; Sequence 18162, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:

;; APPLICANT: Venter, J. Craig
;; APPLICANT: et al.
;; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;; FILE REFERENCE: DROSOPHILA GENES.
;; FILE REFERENCE: CL000728
;; CURRENT APPLICATION NUMBER: US/11/097,143
;; CURRENT FILING DATE: 2005-04-04
;; PRIOR APPLICATION NUMBER: 60/157,832
;; PRIOR FILING DATE: 1999-10-05

;; PRIOR APPLICATION NUMBER: 60/160,191
;; PRIOR FILING DATE: 1999-10-19
;; PRIOR APPLICATION NUMBER: 60/161,932
;; PRIOR FILING DATE: 1999-10-28
;; PRIOR APPLICATION NUMBER: 60/164,769
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/173,383
;; PRIOR FILING DATE: 1999-12-28
;; PRIOR APPLICATION NUMBER: 60/175,693
;; PRIOR FILING DATE: 2000-01-12
;; PRIOR APPLICATION NUMBER: 60/184,831
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: 60/191,637
;; PRIOR FILING DATE: 2000-03-23
;; NUMBER OF SEQ ID NOS: 43008
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 18162
;; LENGTH: 402
;; TYPE: PRT
;; ORGANISM: DROSOPHILA
US-11-097-143-18162

Query Match 54.3%; Score 177; DB 20; Length 402;
Best Local Similarity 73.3%; Pred. No. 4.8e-15;
Matches 33; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 7 DLAGYLLYXPXMRGFGNQADHFLGSLAFKLVRTLAVPPWIEYQ 51
DB 28 DPNGYLYTCFCNGRFGNQADHFLGSLAFKALNRTLILPPWVEYR 72

RESULT 8
US-09-774-954-8
; Sequence 8, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 14:
; . SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-774-954-14

Query Match 24.8%; Score 81; DB 11; Length 28;
Best Local Similarity 94.4%; Pred. No. 0.0021;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMR 20
DB 11 AGSWDPAGYLLYXPXMR 28

RESULT 10
US-10-820-474A-102
; Sequence 102, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
; APPLICANT: LAL, PREETI
; APPLICANT: TANG, Y. TOM
; APPLICANT: GORGONE, GINA A.
; APPLICANT: CORLEY, NEIL C.

US-09-774-954-8

Query Match 35.0%; Score 114; DB 11; Length 474;
Best Local Similarity 64.7%; Pred. NO. 1.6e-06;
Matches 22; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 19 GREGNQADHFLGSLAFKLVRTLAVPPWIEYQH 52
DB 166 GREGNOVDQGLVLAFAKALDRTLVLFPNFIKXH 199

RESULT 9
US-09-774-954-14
; Sequence 14, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 14:
; . SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-774-954-14

Query Match 24.8%; Score 81; DB 11; Length 28;
Best Local Similarity 94.4%; Pred. No. 0.0021;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGSWDLAGYLLYXPXMR 20
DB 11 AGSWDPAGYLLYXPXMR 28

RESULT 10
US-10-820-474A-102
; Sequence 102, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
; APPLICANT: LAL, PREETI
; APPLICANT: TANG, Y. TOM
; APPLICANT: GORGONE, GINA A.
; APPLICANT: CORLEY, NEIL C.

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 61790
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-61790

Query Match 21.3%; Score 69.5; DB 15; Length 198;
Best Local Similarity 50.0%; Pred. No. 0.59;
Matches 16; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

QY 21 FGNQADHFLGSLAFKLVRTLVAV---PPWIE 49
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 FGRQVDSFEGDIAFAGLDVPRVRAVFIRAPWVE 147

RESULT 14
US-10-424-599-266118
; Sequence 266118, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266118
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(152)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_82325C.1.pap
US-10-424-599-266118

Query Match 21.0%; Score 68.5; DB 15; Length 152;
Best Local Similarity 37.2%; Pred. No. 0.61;
Matches 16; Conservative 7; Mismatches 17; Indels 3; Gaps 2;

QY 11 YLLYXPXMGFRGNQADHFLGSLAFKLVRTLVAVPPWIEYQHH 53
::||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 FWYVAPHSQ-FSNQSEFNKNAVIMAGILNRTLVVPPVLD--HH 115

RESULT 15
US-10-437-963-113919
; Sequence 113919, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113919
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17661C.1.pap
US-10-437-963-113919

Query Match 20.7%; Score 67.5; DB 16; Length 538;
Best Local Similarity 38.2%; Pred. No. 3.2;
Matches 21; Conservative 5; Mismatches 24; Indels 5; Gaps 3;

QY 1 RLAGSWDLAG--YLLYXPXMGFRGNQADHFLGSLAFKLVRTLVAVPPWIEYQHH 53
::||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 RCGRSPLEGERFLWYAPHSG-FSNQVGELENAVAALNRTLVVPPVLD--HH 149

Search completed: October 25, 2005, 15:41:23
Job time : 37.8234 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 35.1971 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....LAVPPWIEYQHKKPPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	100.0	388	1	Q9H488 homo sapien
2	353	100.0	388	1	Q6EV69 pan troglod
3	343	97.2	351	2	Q7YRE6 bos taurus
4	343	97.2	391	2	Q7YRE7 bos taurus
5	341	96.6	378	2	Q7YR23 sus scrofa
6	331	93.8	393	1	QFUI MOUSE
7	331	93.8	393	2	Q8C9R4 mus musculus
8	327	92.6	395	2	Q6EV70 rattus norv
9	326	92.4	391	2	Q6EV66 bos taurus
10	308	87.3	380	2	Q8AWB4 gallus gall
11	295	83.3	61	1	QFUI CRIGR
12	287	81.3	395	2	Q7T028 brachydanio
13	284	80.5	380	2	Q640S0 xenopus tro
14	284	80.5	384	2	Q70AG7 fugu rubrip
15	284	80.5	387	2	Q6EV65 tetraodon n
16	284	80.5	396	2	Q6EV71 xenopus tro
17	268	75.9	380	2	Q8AXS8 xenopus lae
18	223	63.2	402	1	QFUI DROME
19	223	63.2	402	2	Q6ESR7 drosophila
20	223	63.2	402	2	Q659S1 drosophila
21	209	59.2	399	2	Q7QHS7 anopheles g
22	207	58.6	381	2	Q6EV67 ciona savig
23	192	54.7	385	2	Q659R9 onchocerca
24	186	52.4	402	2	Q6EV68 onchocerca
25	183	51.8	381	1	Q18014 caenorhabdi
26	177	50.1	434	2	Q659S2 caenorhabdi
27	164	46.5	364	2	Q8MSR1 drosophila
28	161.5	45.8	353	2	Q659S0 bombyx mori
29	79.5	22.5	832	2	Q7YTA5 leishmania
30	78	22.1	354	1	QPSI ASTFA
31	77.5	22.0	538	2	Q7XC59 oryza sativ

check the date

RESULT 1

ID	QFUI_HUMAN	STANDARD;	PRT;	388 AA.
AC	Q9H488; Q14685; Q9BW76;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)			
DE	(Peptide O-fucosyltransferase) (O-FUCT-1).			
GN	Name=POFUT1; Synonyms=K1AA0180;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RL	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.			
RC	TISSUE=Heart;			
RX	MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;			
RA	Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,			
RA	Haltiwanger R.S.;			
RT	"Modification of epidermal growth factor-like repeats with O-fucose:			
RT	molecular cloning and expression of a novel GDP-fucose protein O-			
RT	fucosyltransferase.";			
RL	J. Biol. Chem. 276:40338-40345(2001).			
RL	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Brain;			
RX	MEDLINE=22158633; PubMed=12168954;			
RA	Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;			
RT	"Construction of expression-ready cDNA clones for KIAA genes: manual			
RT	curation of 330 KIAA cDNA clones.";			
RL	DNA Res. 9:99-106(2002).			
RL	[3]			
RP	SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).			
RC	TISSUE=Bone marrow;			
RX	MEDLINE=96281124; PubMed=8724849;			
RA	Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. V.			
RT	The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by			
RT	analysis of cDNA clones from human cell line KG-1.";			
RL	DNA Res. 3:17-24(1996).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;			
RA	Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D.J., Burrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			

Q9AUT5 oryza sativ
Q9L0H6 streptomyce
Q82DG5 streptomyce
Q9IA35 poephila gu
Q9HGR6 mucor circi
Q9I973 serinus can
Q9W6A5 brachydanio
Q8AYM9 brachydanio
Q9OWD5 hippoglossu
Q8GGR3 hippoglossu
Q80IU8 brachydanio
Q8AYM8 brachydanio
Q96ZE4 leishmania
Q9P2K1 homo sapien

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaealiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh L.L., Martin S.L., McComachie I.J., McLeay K., Murray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.B., Soderlund C., Steward C.A., Sulston J.B.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin I.B., Ioshizuka S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gay L.J., Rulyk S.W.,
RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=97175972; PubMed=9023546;
RA Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT enzymatic addition of O-linked fucose to EGF domains."
RL Glycobiology 6:837-842(1996)
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains.
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese.
CC -!- PATHWAY: Glycosylation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9H488-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9H488-2; Sequence=VSP_001809;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC lung, liver, skeletal muscle, kidney and pancreas.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF375884; AAL09576.1; --
CC EMBL; DB0002; BAA11497.2; --
CC EMBL; AL121897; CAC16424.1; --
CC EMBL; BC000582; AAH00582.1; --
CC Genew; HGNC:14948; POSUTL.
CC H-InvDB; HIX0015724; --
CC MTM; 607481; --
CC GO; GO:0005737; C:cytoplasm; ISS.
CC GO; GO:0030173; C:integral to Golgi membrane; NAS.
CC GO; GO:0046922; P:peptide-O-fucosyltransferase activity; TAS.
CC GO; GO:0006922; P:embryonic development; NAS.
CC GO; GO:0003750; P:Notch signaling pathway; ISS.
CC GO; GO:0007219; P:O-glycan processing; TAS.
CC GO; GO:0016266; P:regulation of transcription; NAS.
CC GO; GO:0045449; P:regulation of translation; NAS.
CC GO; GO:0006445; P:regulation of metabolism; Glycoprotein;
CC Alternative splicing; Fucose metabolism; Glycoprotein;
CC Glycosyltransferase; Manganese; Signal; Transferase.
CC SIGNAL 1 26 Potential.
CC CHAIN 27 388
CC GDP-fucose protein O-fucosyltransferase
CC FT 1.
CC FT CARBOHYD 62 62 N-linked (GlcNAc...) (Potential).
CC FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).
CC FT VARSPLIC 182 388
CC FT EQAIHAHLVVRPVGVGIHLRIGSDWKNACAMLKDGAGHFMG
CC SPTCVYSRSTAAPIPTMTCLPDLKEIQRVAVKLWVRLDAQ
CC SVTVATDSYVPELOQLFKGVKVKVSLKPEVAQVLDYILG
CC QADHFTGNCVSSFTAFVKRERDLQGRSSFFGMDRPPKLRD
CC FT EP -> RENHSCVTLPER (in isoform 2).
CC FT /FTId=VSP_001809.
CC FT SEQUENCE 388 AA; 43955 MW; 3FACCCCA434D02415 CRC64;
CC SQ
CC Query Match 100.0%; Score 353; DB 1; Length 388;
CC Best Local Similarity 100.0%; Pred. No. 4.5e-36;
CC Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Qy 1 MPAGSWDPAGVLLYCPGMRFGNQADHFLGSLAFKLNRLTFLAVPFWIEYOHKPPFTNL 60
CC Db 24 MPAGSWDPAGVLLYCPGMRFGNQADHFLGSLAFKLNRLTFLAVPFWIEYOHKPPFTNL 83
CC
CC Qy 61 H 61
CC Db 84 H 84
CC
CC RESULT 2
CC ID_PANTR STANDARD; PRT; 388 AA.
CC ID_OFUL_PANTR STANDARD; PRT; 388 AA.
CC OSEV69;
CC DT 25-OCT-2004 (Rel. 45, Created)
CC DT 25-OCT-2004 (Rel. 45, Last sequence update)
CC DT 25-OCT-2004 (Rel. 45, Last annotation update)
CC DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
CC DE (Peptide fucosyltransferase) (O-FUCT-1).
CC GN Names=FOFUT1; Synonyms=FUT12;
CC OS Pan troglodytes (Chimpanzee).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
CC OK NCBI_TaxID=9598;
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP PubMed=12966037; DOI=10.1093/glycob/cwgl13;
CC RX Martinez-Duncker I., Mollicone R., Breton C.,
CC RA Oriol R.;
CC RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
CC fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
CC RT identification of conserved peptide motifs."
CC RL Glycobiology 13:1C-5C(2003).
CC CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains (By similarity).
CC CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.

CC -|- COPACTOR: Manganese (By similarity).
 CC -|- PATHWAY: Glycosylation.
 CC -|- SIMILARITY: Belongs to the glycosyltransferase 68 family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to a license@isb-sib.ch).

CC EMBL: AJ781500; CAH03712.1; -;
 KW Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
 KW Signal; Transferase. 26 Potential.
 FT SIGNAL 1 GDP-fucose protein O-fucosyltransferase
 FT CHAIN 27 388 1.
 FT
 FT CARBOHYD 62 62 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 160 160 N-linked (GlcNAc...) (Potential).
 FT SEQUENCE 388 AA; 43955 MW; 3FACCA434D02415 CRC64;

Query Match 100.0%; Score 353; DB 1; Length 388;
 Best Local Similarity 100.0%; Pred. No. 4.5e-36;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
 :|||
 Db 24 MPAGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 83

QY 61 H 61
 |
 Db 84 H 84

RESULT 3

QYR6 PRELIMINARY; PRT; 351 AA.
 AC QYR6;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Protein O-fucosyltransferase 1b.
 GN Name-pofuti;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Lioriol C., Germot A., Dupuy F., Maftah A.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY344581; AA02333.1; -;
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl...; IEA.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 351 AA; 39576 MW; BF830F61A7296F42 CRC64;

Query Match 97.2%; Score 343; DB 2; Length 351;
 Best Local Similarity 95.1%; Pred. No. 7.5e-35;
 Matches 58; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
 :|||
 Db 27 LPVGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 86

QY 61 H 61
 |
 Db 87 H 87

RESULT 4
QYR7

ID QYR7 PRELIMINARY; PRT; 391 AA.
 AC QYR7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Protein O-fucosyltransferase 1a.
 GN Name-pofuti;
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Lioriol C., Germot A., Dupuy F., Maftah A.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY344580; AA02332.1; -;
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl...; IEA.
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 391 AA; 44307 MW; 96762AB81A2027AD CRC64;

Query Match 97.2%; Score 343; DB 2; Length 391;
 Best Local Similarity 95.1%; Pred. No. 8.4e-35;
 Matches 58; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
 :|||
 Db 27 LPVGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 86

QY 61 H 61
 |
 Db 87 H 87

RESULT 5

QYR23 PRELIMINARY; PRT; 378 AA.
 AC QYR23;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Protein O-fucosyltransferase (fragment).
 GN Name-POFUT1;
 OS Sus scrofa (pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
 RA "A new superfamily of protein-O-fucosyltransferases, alpha2-
 RA fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
 RT identification of conserved peptide motifs.";
 RT Glycobiology 13:1c-5c(2003).
 DR EMBL: AJ567917; CAD99201.1; -;
 DR GO: GO:0016757; F:transferase activity, transferring glycosyl...; IEA.
 KW Glycosyltransferase; Transferase.
 FT NON_TER 1 1
 FT NON_TER 378 378
 SQ SEQUENCE 378 AA; 42842 MW; 1870DEB0C50713F1 CRC64;

Query Match 96.6%; Score 341; DB 2; Length 378;
 Best Local Similarity 96.7%; Pred. No. 1.5e-34;
 Matches 59; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
 :|||
 Db 20 LPVGSWDPAGYLLYPCMGFRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 79

QY 61 H 61
 |
 Db 80 H 80

```
RESULT 6
OFU1 MOUSE
ID OFU1 MOUSE STANDARD; PRT; 393 AA.
AC Q91ZM2;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE (Peptide O-fucosyltransferase) (O-FucT-1).
GN Name=Pofut1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Liver;
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107949200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RA "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase."
RL J. Biol. Chem. 276:40338-40345 (2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bobak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Murthy D.K., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains (By similarity).
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese (By similarity).
CC -!- PATHWAY: Glycosylation.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF375885; AAL05977.1; -;
DR EMBL; BC046295; AAH46295.1; -;
DR MGD; MGI:2153207; Pofut1.
DR GO; GO:0030173; C:integral to Golgi membrane; ISS.
DR GO; GO:0008417; F:fucosyltransferase activity; ISS.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0007507; P:heart development; IMP.
DR GO; GO:0007399; P:neurogenesis; IMP.
DR GO; GO:0007219; P:Notch signaling pathway; IMP.
DR GO; GO:0006493; P:O-linked glycosylation; ISS.
DR GO; GO:0001756; P:somitogenesis; IMP.
KW Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW SIGNAL; Transferase.
FT SIGNAL 1 30 Potential.
FT CHAIN 31 393 GDP-fucose protein O-fucosyltransferase
FT CASEOHVD 67 67 N-linked (GlcNAc...) (Potential).
FT CASEOHVD 165 165 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 393 AA; 44688 MW; D982104E95E5CF3B CRC64;
Query Match 93.8%; Score 331; DB 1; Length 393;
Best Local Similarity 98.3%; Pred. No. 2.8e+33;
Matches 58; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 AGSWDPAGYLLYPCMGFRGNGQADHFLGSLAFKLNRLTLAVPPWIEYQHKKPFTNLH 61
DB 31 AGSWDLAGYLLYPCMGFRGNGQADHFLGSLAFKLNRLTLAVPPWIEYQHKKPFTNLH 89
RESULT 7
Q8C8R4 PRELIMINARY; PRT; 393 AA.
AC Q8C8R4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE clone: A930028F21 product: protein O-fucosyltransferase 1, full insert
DE sequence.
GN Name=Pofut1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muratsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630 (2000).
RN [5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
```


ID	Genes	Accession	Entry
AC	Q6EV65;	(TrEMBLrel. 28, Created)	
DT	25-OCT-2004	(TrEMBLrel. 28, Last sequence update)	
DT	25-OCT-2004	(TrEMBLrel. 28, Last annotation update)	
DT	25-OCT-2004	(TrEMBLrel. 28, Last annotation update)	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:08:04 ; Search time 1382.21 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool.h/US09774954/runat 25102005.105432 6389/app query.fasta_1.917
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.1 4968 @runat 25102005.105432 6389 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	1100	6	AR163457 Sequence
2	353	100.0	1100	6	BD103196 O-fucosyl
3	353	100.0	1167	6	C0727777 Sequence
4	353	100.0	1167	9	AJ781500 Pan trogl

5	353	100.0	1300	6	AR163452 Sequence
6	353	100.0	1300	6	BD103189 O-fucosyl
7	353	100.0	1514	6	AR163451 Sequence
8	353	100.0	1514	6	BD103188 O-fucosyl
9	353	100.0	1524	9	EC000582 Homo sapi
10	353	100.0	4560	9	AL832023 Homo sapi
11	353	100.0	5189	9	D80002 Homo sapien
12	353	100.0	5249	9	AF375884 Homo sapi
13	353	100.0	5266	6	AX780153 Sequence
14	353	100.0	11284	6	BD103190 O-fucosyl
15	343	97.2	1056	4	AY344581 Bos tauru
16	343	97.2	1176	4	AY344580 Bos tauru
17	341	96.6	1134	4	SSC567917 Sus scro
18	331	93.8	1302	10	AF375885 Mus muscu
19	331	93.8	1529	10	BC046295 Mus muscu
20	327	92.6	1188	10	AJ781499 Rattus no
21	326	92.4	1176	4	AJ781503 Bos tauru
22	308	87.3	920	5	AJ719585 Gallus ga
23	308	87.3	1732	5	AJ535754 Gallus ga
24	308	87.3	2546	5	AJ720352 Gallus ga
25	287	81.3	1549	5	DR579536 Danio rer
26	284	80.5	1155	5	AJ606070 Fugu rubr
27	284	80.5	1164	5	AJ781504 Tetraodon
28	284	80.5	1502	5	BC082519 Xenopus t
29	284	80.5	1538	5	AJ781498 Siurana
30	268	75.9	1505	5	AL514425 Xenopus l
31	233	66.0	145414	9	HLA392M18 Human DNA
32	233	66.0	200329	10	AC078911 Mus muscu
33	233	66.0	208614	10	AL807380 Mouse DNA
34	233	66.0	240336	2	AC094844 Rattus no
35	233	66.0	240510	2	AC134371 Rattus no
36	223	63.2	1209	3	AJ831490 Drosophil
37	223	63.2	1209	6	CQ590403 Sequence
38	223	63.2	1295	3	AV118651 Drosophil
39	223	63.2	1320	6	AR509393 Sequence
40	223	63.2	1579	3	AB093572 Drosophil
41	218	61.8	5009	6	AR163453 Sequence
42	218	61.8	5009	6	BD103191 O-fucosyl
43	207	58.6	1146	3	AJ781502 Ciona sav
44	200	56.7	1841	3	AK112708 Ciona int
45	192	54.4	1160	3	AJ831492 Onchocerc

ALIGNMENTS

RESULT 1	AR163457	Sequence 16 from patent US 6270987.	1100 bp	DNA	lin
LOCUS	AR163457	Sequence 16 from patent US 6270987.	1100 bp	DNA	
DEFINITION	AR163457	Sequence 16 from patent US 6270987.	1100 bp	DNA	
ACCESSION	AR163457	Sequence 16 from patent US 6270987.	1100 bp	DNA	
VERSION	AR163457.1	GI:16234062	1100 bp	DNA	
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1100)				
AUTHORS	Wang, Y. and Spellman, M.W.				
TITLE	O-fucosyltransferase				
JOURNAL	Patent: US 6270987-A 16 07-AUG-2001;				
FEATURES	Location/Qualifiers				
source	1..1100				
	/organism="unknown"				
	/mol_type="unassigned DNA"				

ORIGIN

Alignment Scores:					
Pred. No.:	1.12e-35	Length:	1100		
Score:	353.00	Matches:	61		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		

US-09-774-954-9 (1-61) x ARI63457 (1-1100)

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 Db 1 ATGCCCGGGGCTCTGGAGCCCGCGGTACCTGCTCTACTGCCCTGCATGGGGCGC 60
 QY 21 PheGlyAenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 Db 61 TTTGGGACCAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120
 QY 41 ThrLeuAlaValProProTPIleGluTyrGlnHisLysProProPheThrAsnLeu 60
 Db 121 ACCTGGCTGCTCCCTCTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 180
 QY 61 His 61
 Db 181 CAT 183

RESULT 2
 BDI03196
 LOCUS BDI03196 1100 bp DNA linear PAT 27-AUG-2002
 DEFINITION O-fucosyltransferase.
 ACCESSION BDI03196
 VERSION BDI03196.1 GI:22648770
 KEYWORDS JP 2001527389-A/9.
 SOURCE unidentified
 ORGANISM unidentified
 unclassified.

REFERENCE 1 (bases 1 to 1100)
 AUTHORS Wang, Y. and Spellman, M.W.
 TITLE O-fucosyltransferase
 JOURNAL GENETECH INC
 COMMENT OS Unidentified
 PN JP 2001527389-A/9
 PD 25-DEC-2001
 PF 17-DEC-1997 JP 1998532877
 PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI

PC C12N15/54, C12N9/10, C07K16/40
 CC Strandedness: Single;
 CC Topology: Linear;
 CC O-fucosyltransferase

FT Key Location/Qualifiers
 FT source 1..1100
 FT /organism='Unidentified'.
 FT Location/Qualifiers
 1..1100
 /organism='unidentified'
 /mol_type='genomic DNA'
 /db_xref='taxon:32644'

FEATURES
 source

ORIGIN

Alignment Scores:
 Pred. No.: 1.12e-35 Length: 1100
 Score: 353.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x BDI03196 (1-1100)
 QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 Db 1 ATGCCCGGGGCTCTGGAGCCCGCGGTACCTGCTCTACTGCCCTGCATGGGGCGC 60
 QY 21 PheGlyAenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 Db 61 TTTGGGACCAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120
 QY 41 ThrLeuAlaValProProTPIleGluTyrGlnHisLysProProPheThrAsnLeu 60

Alignment Scores:
 Pred. No.: 1.12e-35 Length: 1100
 Score: 353.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x BDI03196 (1-1100)
 QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 Db 1 ATGCCCGGGGCTCTGGAGCCCGCGGTACCTGCTCTACTGCCCTGCATGGGGCGC 60
 QY 21 PheGlyAenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 Db 61 TTTGGGACCAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 120
 QY 41 ThrLeuAlaValProProTPIleGluTyrGlnHisLysProProPheThrAsnLeu 60

Db 121 ACCTGGCTGCTCCCTCTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 180
 QY 61 His 61
 Db 181 CAT 183
 RESULT 3
 BDI03196
 LOCUS BDI03196 1167 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 13711 from Patent WO02068579.
 ACCESSION CQ727777
 VERSION CQ727777.1 GI:42294771
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
 TITLE Kits, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 JOURNAL Patent: WO 02068579-A 13711 06-SEP-2002;
 PE Corporation (NY) (US)

FEATURES
 source

ORIGIN

Alignment Scores:
 Pred. No.: 1.19e-35 Length: 1167
 Score: 353.00 Matches: 61
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x CQ727777 (1-1167)

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 Db 70 ATGCCCGGGGCTCTGGAGCCCGCGGTACCTGCTCTACTGCCCTGCATGGGGCGC 129
 QY 21 PheGlyAenGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 Db 130 TTTGGGACCAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCCAAGCTGCTAAACCGT 189
 QY 41 ThrLeuAlaValProProTPIleGluTyrGlnHisLysProProPheThrAsnLeu 60
 Db 190 ACCTTGGCTGCTCCCTCTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 249

QY 61 His 61
 Db 250 CAT 252

RESULT 4
 BDI03196
 LOCUS BDI03196 1167 bp mRNA linear PRI 07-JUL-2004
 DEFINITION Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
 gene).

ACCESSION AJ781500

VERSION AJ781500.1 GI:50057069

KEYWORDS fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1
 AUTHORS Martinez-Duncker, I., Mollicone, R., Candellier, J.J., Breton, C. and
 Oriol, R.

TITLE A new superfamily of protein-O-fucosyltransferases.

alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
Phylogeny and identification of conserved peptide motifs
Glycobiology 13 (12), 1C-5C (2003)
12966037

2 Martinez-Duncker I., Oriol R. and Mollicone R.
Unpublished

3 (bases 1 to 1167)

REFERENCE
AUTHORS Oriol, R.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
Vaillant-Couturier, 94807, FRANCE

FEATURES
source

1. .1167
/organism="Pan troglodytes"
/mol_type="mRNA"
/db_xref="taxon:9598"

gene
1. .1167
/gene="fut12"

CDS
1. .1167
/gene="fut12"

/function="adds fucose to Ser or Thr"

/notes="pofut1"

/codon_start=1

/product="protein-O-fucosyltransferase 1"

/protein_id="CAH03712.1"

/db_xref="GI:50057070"

/translation="MGAAARPLSVSFLLLPLPGMPAGSWDPAGYLLYPCMGFR
GNAQDFLAGLAFKLNLTAVPPWIEYQHHKPPFTNLHVSQYKFKLEPLQAYHRV
ISLEDFMEKLAHPHPPEKRVAYCEVAAQSPDKKCPMKGNPPGPPWDOFHYSEN
KSEFTGISFASYSQRSPKHPVLALPGAPQFPVLEHRPLQKYMWSDEM
VKTGEAIIHAHLVPVYVGIHLRIGSDWKNACMLKDGTHAGHFMASPCQVGYSTAA
PLTMTMCLPLKEIQRAVLKWLRLDAQSVYVATDSYVPELQQLFKGKVKVSLKP
EVAQVDLYILGQADHFIQNCVSFTAFVXRERDLQGRPSFFGMDRPPKLRDEF"

ORIGIN

Alignment Scores:
Pred. No.: 1.19e-35 Length: 1167
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-9 (1-61) x AJ781500 (1-1167)

QY 1 MetProAlaGlySerTpaAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 70 ATGCCCGCGGGCTCTCTGGGACCGCGGGTTACTGCTCTACTGCCCTGCATGGGGCGC 129
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 130 TTGGGACCGAGCCGATCATCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 189
QY 41 ThrLeuAlaValProProTrrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60
DB 190 ACCTTGCTGTCCTCTCTGGATTGAGTACGAGCATCACAGGCTCCTTTTACCAACCTC 249
QY 61 His 61
DB 250 CAT 252

RESULT 5
AR163452
LOCUS AR163452 1300 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 6 from patent US 6270987.
ACCESSION AR163452
VERSION AR163452.1 GI:162234057
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 6 07-AUG-2001;
FEATURES
source
1. .1300
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.33e-35 Length: 1300
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x AR163452 (1-1300)

QY 1 MetProAlaGlySerTpaAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 136 ATGCCCGCGGGCTCTCTGGGACCGCGGGTTACTGCTCTACTGCCCTGCATGGGGCGC 195
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 196 TTGGGACCGAGCCGATCATCTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 255
QY 41 ThrLeuAlaValProProTrrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60
DB 256 ACCTTGCTGTCCTCTCTGGATTGAGTACGAGCATCACAGGCTCCTTTTACCAACCTC 315
QY 61 His 61
DB 316 CAT 318

RESULT 6

BD103189
LOCUS BD103189 1300 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103189
VERSION BD103189.1 GI:22648763
KEYWORDS JP 2001527389-A/2.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 2 25-DEC-2001;
GENENTECH INC

COMMENT
OS Unidentified
PN JP 2001527389-A/2
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
PC 12N15/54, C12N9/10, C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FH Key Location/Qualifiers
FT source 1. .1300
/organism="Unidentified".

FEATURES
source
1. .1300
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 1.33e-35 Length: 1300

Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x BD103189 (1-1300)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||||
DB 136 ATGCCCGCGGGCTCTGGAGCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 195
|||||

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||||
DB 196 TTTGGGAACAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 255
|||||

QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60
|||||
DB 256 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAGACCTCTCTTTCACCAACCTC 315
|||||

QY 61 His 61
|||||
DB 316 CAT 318

RESULT 7
LOCUS ARI163451
DEFINITION Sequence 2 from patent US 6270987.
ACCESSION ARI163451
VERSION ARI163451.1 GI:16234056
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 2 07-AUG-2001;
FEATURES
source Location/Qualifiers
1..1514
/organism="unassigned DNA"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1.57e-35 Length: 1514
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x ARI163451 (1-1514)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||||
DB 1 ATGCCCGCGGGCTCTGGAGCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60
|||||

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||||
DB 61 TTTGGGAACAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
|||||

QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60
|||||
DB 121 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAGACCTCTCTTTCACCAACCTC 180
|||||

QY 61 His 61
|||||
DB 181 CAT 183

RESULT 8
LOCUS BD103188
DEFINITION Homo sapiens protein O-fucosyltransferase 1, transcript variant 2,
mRNA (cDNA clone MGC:2482 IMAGE:3162533), complete cds.
ACCESSION BC000582
VERSION BC000582.2 GI:37588976
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1524)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

DEFINITION O-fucosyltransferase.
ACCESSION BD103188
VERSION BD103188.1 GI:22648762
KEYWORDS JP 2001527389-A/1.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1514)
AUTHORS Wang, Y. and Spellman, M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 1 25-DEC-2001;
GENENTECH INC
COMMENT CS Unidentified
PN JP 2001527389-A/1
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
PC C12N15/54, C12N9/10, C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FH Key Location/Qualifiers
FT Source 1..1514
/organism="Unidentified".
Location/Qualifiers
1..1514
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 1.57e-35 Length: 1514
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x BD103188 (1-1514)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||||
DB 1 ATGCCCGCGGGCTCTGGAGCCGCGGTACCTGCTCTACTGCCCCCTGCATGGGGCGC 60
|||||

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||||
DB 61 TTTGGGAACAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
|||||

QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60
|||||
DB 121 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAGACCTCTCTTTCACCAACCTC 180
|||||

QY 61 His 61
|||||
DB 181 CAT 183

RESULT 9
LOCUS BC000582
DEFINITION Homo sapiens protein O-fucosyltransferase 1, transcript variant 2,
mRNA (cDNA clone MGC:2482 IMAGE:3162533), complete cds.
ACCESSION BC000582
VERSION BC000582.2 GI:37588976
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1524)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, D., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McGowan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalau, D.E., Schnerch, A., Schein, J.E., Jones, S.D., and Marra, M.A. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1524)

Strausberg, R.

Direct Submission

Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Oct 8, 2003 this sequence version replaced gi:12653608.

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tauregon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 5 Row: 0 Column: 9

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27436890.

Location/Qualifiers

1. 1524

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:2482 IMAGE:3162533"

/tissue_type="Brain, neuroblastoma"

/clone_lib="NIH MGC-19"

/lab_host="DH10B-R"

/note="vector: pOTB7"

1. 1524

/gene="pOFUT1"

/note="synonyms: KIAA0180, O-FUT, O-Fuc-T, FUT12, MGC2482, O-FucT-1"

/db_xref="LocusID:23509"

/db_xref="MIM:607491"

50. .634

/gene="pOFUT1"

/codon_start=1

/product="protein O-fucosyltransferase 1, isoform 2"

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, D., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McGowan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalau, D.E., Schnerch, A., Schein, J.E., Jones, S.D., and Marra, M.A. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 1524)

Strausberg, R.

Direct Submission

Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Oct 8, 2003 this sequence version replaced gi:12653608.

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tauregon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 5 Row: 0 Column: 9

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27436890.

Location/Qualifiers

1. 1524

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:2482 IMAGE:3162533"

/tissue_type="Brain, neuroblastoma"

/clone_lib="NIH MGC-19"

/lab_host="DH10B-R"

/note="vector: pOTB7"

1. 1524

/gene="pOFUT1"

/note="synonyms: KIAA0180, O-FUT, O-Fuc-T, FUT12, MGC2482, O-FucT-1"

/db_xref="LocusID:23509"

/db_xref="MIM:607491"

50. .634

/gene="pOFUT1"

/codon_start=1

/product="protein O-fucosyltransferase 1, isoform 2"

precursor"

/protein_id="AAH00582.1"

/db_xref="GI:12653609"

/db_xref="LocusID:23509"

/db_xref="MIM:607491"

/translation="MGAARARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCWGRFQNAADFLGSLAFKLLNRTLAVPWYQHHKPPFTNLHVSYQKTKLEPLQAYHRVIGLSEDFMEKLAPTHWPPEKRVAYCFEVAARSPDKTKCPMKEGNPFPWDPQHVSGFNKSELFSGISFSASYREQWSQRRENHSCVTLLEPR"

ORIGIN

Alignment Scores:

Pred. No.: 1524

Score: 353.00 Matches: 61

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-09-774-954-9 (1-61) x BC000582 (1-1524)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20

DB 119 ATGCCCTGCGGGCTCTGGGACCGCGCGGTACCTGCTCTACTGCCCTCGATGGGGCGC 178

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40

DB 179 TTGGGAACAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 238

QY 41 ThrLeuAlaValProTrpIleGluTyrGlnHisLysProProPheThrAsnLeu 60

DB 239 ACCTTGGCTGCTCTCTTGGATTGATGACCATCACAAAGCTCTTTTCAACACCTC 298

QY 61 His 61

DB 299 CAT 301

RESULT 10

LOCUS HSM803330 4560 bp mRNA linear PRI 13-MAY-2003

DEFINITION Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).

ACCESSION AL832023

VERSION AL832023.1 GI:21732563

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4560)

AUTHORS Wambutt, R., Heubner, D., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp451J114) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

Location/Qualifiers

1. 4560

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="RZPD:DKFZp451J114"

/db_xref="taxon:9606"

/clone="DKFZp451J114"

/tissue_type="human skeletal muscle"

/clone_lib="451 (synonym: hlcc1). Vector pSport1; host DH10B; sites NotI + SalI"

```

polya_site      /dev_stage="adult"
4538
ORIGIN
Alignment Scores:      Length:      4560
Pred. No.:            5.09e-35      Matches:      61
Score:                353.00        Conservative:  0
Percent Similarity:    100.00%       Mismatches:   0
Best Local Similarity: 100.00%       Indels:       0
Query Match:          100.00%       Gaps:        0
DB:
US-09-774-954-9 (1-61) x HSM803330 (1-4560)
QY      1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db      51 ATGCGTCGGGGCTCTGGGACCGCGGTTACTGCTCTACTGCGGCTGCATGGGGGCG 110
QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      111 TTGGGAACCGCGGATCACTCTTGGGCTCTGGCATTTGCAAGCTGCTAAACCGT 170
QY      41 ThrLeuAlaValProProTTPileGluTyrGlnHisLysProProPheThrAsnLeu 60
Db      171 ACCTTGGCTGCTCCCTTGGATGGATGACGATCAAGGCTCTCTTCCACCAACCTC 230
QY      61 His 61
Db      231 CAT 233
RESULT 11
LOCUS      D80002      Homo sapiens KIAA0180 mRNA, complete cds.
DEFINITION      D80002.2 GI:20521837
ACCESSION      D80002
VERSION
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
source
1. .5189
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="caxon:9606"
/clone="ha02567s1"
/sex="male"

```

```

/cell_line="KG-1"
/tissue_type="myeloblast"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
/notes="This sequence was obtained by subcloning of the DNA
fragments derived from two cDNA clones (1 - 1066 was
derived from pf00933 and 1067 - 5189 was derived from
ha02567)."
1. .5189
/gene="KIAA0180"
<1. .1213
/gene="KIAA0180"
/citation=[2]
/codon_start=2
/protein_id="BAAL1497.2"
/db_xref="GI:20521838"
/translation="PPRLCAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
DPAGVLLYCPQWGRFGNQADHFLGSLAFKLNRTLAVPFWIEYQHHKPPFTNHLVSY
QKYLKLEPLQYHRVLSLEDMEKLAPTHWPPEKRVAYCFEVAARSPDKKCPKMEG
NPFGLPQHFVSNKSELFTGISFSASVQSFQSPKEHPVLALPGAPQFPVLE
EHRPLQKYMVMSDEMVKTEGAIHAHLVRPVGIIHLAGISGDKWKNACMLKDSGTAGSHF
MASPOCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVVYVATDSYVPEL
COLPKGVKVSLSLKPEVAQVDLYILQADHFTGNCVSSTAFVKRERDLQGRPSFPFG
MDRPFKLKDEF"
ORIGIN
Alignment Scores:      Length:      5189
Pred. No.:            5.85e-35      Matches:      61
Score:                353.00        Conservative:  0
Percent Similarity:    100.00%       Mismatches:   0
Best Local Similarity: 100.00%       Indels:       0
Query Match:          100.00%       Gaps:        0
DB:
US-09-774-954-9 (1-61) x D80002 (1-5189)
QY      1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db      116 ATGCGTCGCGGGCTCTGGGACCGCGGTTACTGCTCTACTGCGGCTGCATGGGGCGC 175
QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      176 TTGGGAACCGCGGATCACTCTTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGT 235
QY      41 ThrLeuAlaValProProTTPileGluTyrGlnHisLysProProPheThrAsnLeu 60
Db      236 ACCTTGGCTGCTCCCTTGGATGGATGACGATCAAGGCTCTCTTCCACCAACCTC 295
QY      61 His 61
Db      296 CAT 298
RESULT 12
AF375884
LOCUS
DEFINITION      Homo sapiens protein o-fucosyltransferase (FOFUT1) mRNA, complete
ACCESSION      AF375884
VERSION
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED

```

REFERENCE 2 (bases 1 to 5249)
AUTHORS Shi,S.; Stanley,P.; Wang,Y.; Shao,L.; Harris,R.J.; Spellman,M.W.
and Haltiwanger,R.S.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES
source
1. .5249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="between PLAGL2 and KIF3B"
/tissue_type="heart"
1. .5249
/genes="POPUT1"
50. .1216
/genes="POPUT1"
/function="catalyzes transfer of fucose from GDP-fucose to
serine or threonine on epidermal growth factor-like
repeats"
/notes="O-FucT-1; glycosyltransferase"
/codon_start=1
/product="protein O-fucosyltransferase"
/protein_id="AAI09576.1"
/db_xref="GI:15825114"
/translation="MGAAWAARPLSVSFLILLPLPMPAGSWDPAGYLLYCPMGGRF
GNQADHFLGSLAFLKLNRLTAVPPWIEYQHKKLPFTNLHVSYQKFKLEPLQAYHRV
ISLEDFWEKLAFTHTWKEPVAYCFEVAQORSDDKTCPMKEGNPFQWQPHVSEFN
KSLFTG1SFSASVREQWSPFKHEPVLALPCAPQFPVLEHRLPLQKYMVMSDEM
VKTGEAQLHAHLVPVYVGIHLRIGSDWKNACMLKDGTAGHFMASPOCVGYSRSTAA
PLTWTCPLDLKSIQRAVKLWVSLDAQSVYVATDSYVPELQQLPKGKVKVLSLKP
EVAQVDLIILGQADHFIQNCVSSFTAPVKRERDLQGRPSSFPGMRPPKLDRDF"

Alignment Scores:
Pred. No.: 5,92e-35 Length: 5249
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-9 (1-61) x AF375884 (1-5249)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
119 ATGCCCGGGGCTCTCTGGAGCCGGCTTACCTGCTCTACTGCCCTGCATGGGGCGC 178

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
179 TTGTGGGAACAGGCGCATCTCTTGGGCTCTCTGGCATTTGCANAAGCTGTAAACCGT 238

QY 41 ThrLeuAlaValProProTyrPilleGluTyrGlnHisLysProProPheThrAsnLeu 60
239 ACCTTGGCTGTCTCTCTTGGATTGATGACCAAGCATCACAGGCTCTCTTACCACCACTC 298

QY 61 His 61
299 CAT 301

RESULT 13
AX780153
LOCUS AX780153 5266 bp DNA linear PAT 14-JUL-2003
DEFINITION Sequence 2310 from Patent WO03039443.
ACCESSION AX780153
VERSION AX780153.1 GI:32697147.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1

AUTHORS Haerlach,T.; Schoch,C.; Kern,W.; Kohlmann,A.; Schnittger,S.;
Dugas,M.; Eils,R.; Brors,B. and Mergenthaler,S.
TITLE Novel genetic markers for Leukemias
JOURNAL Patent: WO 03039443-A 2310 15-MAY-2003;
Deutsches Krebsforschungszentrum (DE);
Ludwig-Maximilian-Universitaet Muenchen (DE);
PD Dr. Dr. Schoch, Claudia (DE); Kern, Wolfgang (DE)
FEATURES
source
1. .5266
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 5,94e-35 Length: 5266
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x AX780153 (1-5266)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
132 ATGCCCTGGGGCTCTCTGGAGCCGGCTTACCTGCTCTACTGCCCTGCATGGGGCGC 191

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
192 TTGTGGGAACAGGCGCATCTCTTGGGCTCTCTGGCATTTGCANAAGCTGTAAACCGT 251

QY 41 ThrLeuAlaValProProTyrPilleGluTyrGlnHisLysProProPheThrAsnLeu 60
252 ACCTTGGCTGTCTCTCTTGGATTGATGACCAAGCATCACAGGCTCTCTTACCACCACTC 311

QY 61 His 61
312 CAT 314

RESULT 14
BD103190
LOCUS BD103190 11284 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103190
VERSION BD103190.1 GI:22648764
KEYWORDS JP 2001527389-A/3.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 11284)
AUTHORS Wang,J. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 3 25-DEC-2001;
GENENTECH INC
COMMENT OS Unidentified
PN JP 2001527389-A/3
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498, 26-NOV-1997 US 08/978741 PI
YANG WANG, MICHAEL W SPELLMAN
PC C12N15/54, C12N9/10, C07K16/40
CC Strandedness: Single;
CC Topology: linear;
CC O-fucosyltransferase
FH Key
FT source
FT Location/Qualifiers
1. .11284
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores: 1.34e-34 Length: 11284
Pred. No.: 353.00 Matches: 61
Score: 353.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x BD103190 (1-11284)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 4236 ATGCCCGCGGGCTCTGGACCGCGCGGTACTGCTCTACTGCCCCGTCATGGGCGC 4295

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 4296 TTTGGGAACCGAGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 4355

QY 41 ThrLeuAlaValProProTrrPileGluTyrGlnHisLysProPheThrAsnLeu 60
DB 4356 ACCTTGGCTGTCCTCCCTTGATTTAGTACCAAGCATCAAGCCCTCTTTTACCACCTC 4415

QY 61 His 61
DB 4416 CAT 4418

RESULT 15
LOCUS AY344581 1056 bp mRNA linear MAM 03-AUG-2003
DEFINITION Bos taurus protein O-fucosyltransferase 1b (pofut1) mRNA, complete cds.
ACCESSION AY344581
VERSION AY344581.1 GI:33303531
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 1056)
AUTHORS Lioriol,C., Germot,A., Dupuy,F. and Maftah,A.
TITLE Genomic organization and expression profile of O-fucosyltransferase genes, pofut1 and pofut2, in Bos taurus
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1056)
Lioriol,C.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie, Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges 87060, France

FEATURES

source Location/Qualifiers
1..1056
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
1..1056
/gene="pofut1"
1..1056
/gene="pofut1"
/note="O-FuCT-1b"
/codon_start=1
/product="protein O-fucosyltransferase 1b"
/protein_id="AA020333.1"
/db_xref="GI:33303532"
/translation="MGAAAWAPSLPPRVSLLLLLLLPLPGLPVGSDPDAGYLLYCPCM GRFGNADHFLGSLAFAPKLNLRLTAPWIEYQHHKPEFTNVHVSQKYPKLEPLQAY HRVILEDFMEKLAFTWPEKRVAYCFEVAQRSPOKCTCFMKENPFPGFDQFHV SNKSELFAGISFSASYKQDWIQRFSPBEHPVLALPGAPQFPVLEHRPLQKYMWS DENVTRTGAQIHAHLIRPYVGIHLIRIGSDWNACAMKDGTAGAHFMAWSPQCVGYSRH TTAPLTMTMCLPDLKLEIRRALKLWVTALNAQSVYIATDSSEYLPETIQOQLFKGKSLSPA VPKASEMSLLLGDRKGR"

gene

CDS

ORIGIN

Alignment Scores: 2.07e-34 Length: 1056
Pred. No.: 343.00 Matches: 58
Score: 343.00
Percent Similarity: 98.36% Conservative: 2
Best Local Similarity: 95.08% Mismatches: 1
Query Match: 97.17% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-9 (1-61) x AY344581 (1-1056)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 79 CTGCCGTGGGCTCTCTGGGATCCGGCCGGTTACTGCTCTACTGCCCCATGCATGGGCGT 138

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 139 TTTGGGAACCGAGCTGATCATTTCTTGGGCTCCCTGGCATTGCAAGCTGTTGAACCGC 198

QY 41 ThrLeuAlaValProProTrrPileGluTyrGlnHisLysProPheThrAsnLeu 60
DB 199 ACCTTGGCTGATCCCTTGATTTAGTACCAAGCATCAAGCCCTCTTTTACCACCTC 258

QY 61 His 61
DB 259 CAT 261

Search completed: October 26, 2005, 12:55:33
Job time : 1387.21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 05:59:38 ; Search time 1065.43 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGK.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat 25102005 105432 6406/app query.fasta_1.917
-DB=EST -QMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09774954 @CGN 1.1 5334 @runat 25102005 105432 6406 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gasl.*
9: gb_gasl2.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	353	100.0	520	7 CR543310	CR543310 DKFZp459B
2	353	100.0	550	7 CN348927	CN348927 170005999
3	353	100.0	551	2 BE261136	BE261136 601152308
4	353	100.0	576	2 BE262956	BE262956 601147417
5	353	100.0	595	7 CN348926	CN348926 170005326
6	353	100.0	610	1 AL555719	AL555719 AL555719
7	353	100.0	656	7 CV029252	CV029252 7968 Full1
8	353	100.0	699	7 AU137536	AU137536 AU137536
9	353	100.0	705	7 CN348925	CN348925 170005328

10	353	100.0	753	4 BG424731	BG424731 602453575
11	353	100.0	790	4 BG519639	BG519639 602578756
12	353	100.0	808	4 BG747760	BG747760 602705255
13	353	100.0	841	5 BX345989	BX345989 BX345989
14	353	100.0	889	4 B1524031	B1524031 603052202
15	353	100.0	898	2 BF315757	BF315757 601900828
16	353	100.0	919	5 BUI55142	BUI55142 AGENCOURT
17	353	100.0	942	7 CO774833	CO774833 ILLUMITGEN
18	353	100.0	950	4 B1199190	B1199190 602758739
19	353	100.0	955	2 BE260030	BE260030 601150556
20	353	100.0	1555	3 CR617950	CR617950 full-length
21	353	100.0	1711	2 BF206548	BF206548 601869779
22	350	99.2	624	7 CN348922	CN348922 170004249
23	344	97.5	898	5 BX449172	BX449172 BX449172
24	343	97.2	417	2 AW415368	AW415368 49609 MAR
25	341	96.6	473	2 BP444498	BP444498 262391 MA
26	341	96.6	666	5 BP459416	BP459416 BP459416
27	341	96.6	736	5 BP457865	BP457865 BP457865
28	340	96.3	729	2 BE272703	BE272703 601105080
29	338	95.8	689	4 B1559635	B1559635 603252494
30	335	94.9	1241	4 BM478111	BM478111 AGENCOURT
31	333	94.3	874	4 B1757540	B1757540 603029693
32	331	93.8	318	6 BY778150	BY778150 BY778150
33	331	93.8	352	5 BY328132	BY328132 BY328132
34	331	93.8	456	2 BB633858	BB633858 BB633858
35	331	93.8	480	2 BB851550	BB851550 BB851550
36	331	93.8	536	7 CF155036	CF155036 B0620C09-
37	331	93.8	604	2 BB618795	BB618795 BB618795
38	331	93.8	605	6 BY728018	BY728018 BY728018
39	331	93.8	611	2 BB623905	BB623905 BB623905
40	331	93.8	614	6 BY752500	BY752500 BY752500
41	331	93.8	634	2 BB631131	BB631131 BB631131
42	331	93.8	641	2 BB611933	BB611933 BB611933
43	331	93.8	650	6 BY752497	BY752497 BY752497
44	331	93.8	661	2 BB642654	BB642654 BB642654
45	331	93.8	664	6 BY728337	BY728337 BY728337

ALIGNMENTS

RESULT 1
CR543310
LOCUS CR543310 520 bp mRNA linear EST 07-JUL-2004
DEFINITION DKFZp459B068_r1 459 (synonym: pcor1) Pongo pygmaeus CDNA clone
ACCESSION DKFZp459B068_5', mRNA sequence.
VERSION CR543310
KEYWORDS CR543310.1 GI:49894593
SOURCE EST.
ORGANISM Pongo pygmaeus (orangutan)
Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Theria; Primates; Catarrhini; Hominoidea; Pongo.
REFERENCE 1 (bases 1 to 520)
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Pongo, G., Han, M., and Wiemann, S.
TITLE Pongo pygmaeus mRNA (Koehrer, K., Beyer, A., Mewes, H.W., et al.)
JOURNAL Unpublished (2004)
COMMENT Contact: MIPS
MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert clone from S. Wiemann,
Molecular Genome Analysis, German Cancer Research Center (DKFZ);
Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical
Research Center at the Heinrich-Heine-University,
Duesseldorf/Germany) within the CDNA sequencing consortium of the
the RZPD (German Genome Project). This clone (DKFZp459B068) is available at
Heuberweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
clone@rzd.de. Further information about the clone and the
sequencing project is available at
http://mips.gsf.de/projects/cdna/.
FEATURES
source 1..520
Location/Qualifiers

/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp4598068"
/tissue_type="cortex"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="459 (synonym: pcorl)"
/note="Vector: pSport1_Sfi; Site_1: SfiI; Site_2: SfiIb"

ORIGIN

Alignment Scores:
Pred. No.: 1.52e-33 Length: 520
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-774-954-9 (1-61) x CR543310 (1-520)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
116 ATGCCCCGGGCTCTGGACCCGGCGGTACCTGCTCTACTGCCCTGTCATGGGGCGC 175
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
176 TTGGGAACCGAGGGATCACTTCTTGGGCTCTTTGGCATTTGCAAGCTGCTAAACCGT 235
QY 41 ThrLeuAlaValProTrpTyrGlnHisLysProProPheThrAsnLeu 60
236 ACCTTGCTGCTCCCTCTGGATTGATGACGATCACAAGCTGCTTCAACACCTC 295
QY 61 His 61
DB 296 CAT 298

RESULT 2
CN348927
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CN348927 550 bp mRNA linear EST 16-MAY-2004
17000599936929 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
CN348927
EST.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 550)
Brandenberger R., Wei H., Zhang S., Lei S., Murage J., Fisk G.J.,
Li Y., Xu C., Rang R., Guegler K., Rao M.S., Mandalam R.,
Lebkowski J. and Stanton L.W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: tbrandenberger@geron.com
Insert Length: 550 Sfd Error: 0.00.
Location/Qualifiers

FEATURES

source

1. .550
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and
mitogen-treated hES cell line H7"
/clone_lib="GRN_PRENEU"
/note="oligo dt primed, full-length enriched cDNA library
from hES cell line H7 (p29) maintained in feeder-free

conditions. Embryoid bodies were generated in the presence
of all-trans retinoic and mitogens."

ORIGIN

Alignment Scores:
Pred. No.: 1.63e-33 Length: 550
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-774-954-9 (1-61) x CN348927 (1-550)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
109 ATGCTGCGGGGCTCTGGACCCGGCGGTACTCTCTACTGCCCTGTCATGGGGCGC 168
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
169 TTGGGAACCGAGCGGATCACTTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 228
QY 41 ThrLeuAlaValProTrpTyrGlnHisLysProProPheThrAsnLeu 60
229 ACCTTGCTGCTCCCTCTGGATTGATGACGATCACAAGCTGCTTCAACACCTC 288
QY 61 His 61
DB 289 CAT 291

RESULT 3
BE261136
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE261136 551 bp mRNA linear EST 26-OCT-2000
601152308F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508464 5',
mRNA sequence.
BE261136
EST.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 551)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LNCM189 row: P column: 01
High quality sequence stop: 551.
Location/Qualifiers

FEATURES

source

1. .551
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3508464"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NTH MGC 19"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.: 1.64e-33 Length: 551
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-9 (1-61) x BE261136 (1-551)

QY 1 MetProAlaGlySerTirpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||||
Db 100 ATGCGTGGGGCTCTCTGGGACCGCGGTTACTCTCTACTGCGCCCTGATGGGGCGC 159
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||||
Db 160 TTGGGAACAGCCGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 219
QY 41 ThrLeuAlaValProProTirpLeuGlyTyrGlnHisLysProProPheThrAsnLeu 60
|||||
Db 220 ACCTTGGCTGCTCTCTGGATTTAGTACAGCATCACAAAGCTCTCTTACCAACCTC 279
QY 61 His 61
|||
Db 280 CAT 282

RESULT 4

BE262956 576 bp mRNA linear EST 26-OCT-2000
LOCUS 601147171 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3162533 5',
DEFINITION mRNA sequence.
ACCESSION BE262956
VERSION BE262956.1 GI:9136425
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 576)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM122 row: b column: 06
High quality sequence start: 11
High quality sequence stop: 576.

FEATURES

source
1. 576
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3162533"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_19"
/note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1.73e-33 Length: 576
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-9 (1-61) x BE262956 (1-576)

QY 1 MetProAlaGlySerTirpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
|||||
Db 154 ATGCGTGGGGCTCTCTGGGACCGCGGTTACTCTCTACTGCGCCCTGATGGGGCGC 213
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
|||||
Db 214 TTGGGAACAGCCGATCCTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 273
QY 41 ThrLeuAlaValProProTirpLeuGlyTyrGlnHisLysProProPheThrAsnLeu 60
|||||
Db 274 ACCTTGGCTGCTCTCTGGATTTAGTACAGCATCACAAAGCTCTCTTACCAACCTC 333
QY 61 His 61
|||
Db 334 CAT 336

RESULT 5

CN348926 595 bp mRNA linear EST 16-MAY-2004
LOCUS 17000532646873 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
DEFINITION CN348926
ACCESSION CN348926
VERSION CN348926.1 GI:47348860
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 595)
TITLE Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
Li, Y., Xu, C., Fang, R., Guegier, K., Rao, M.S., Mandalam, R.,
Lebkowski, J. and Stanton, L.W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert length: 595 Std Error: 0.00.

FEATURES

source
1. 595
/location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, cell lines H1, H7, and
H9"
/clone_lib="GRN ES"
/note="oligo dT primed, full-length enriched cDNA library
from undifferentiated hES cell lines H1 (p32), H7 (p29),
and H9 (p26) maintained in feeder-free conditions"

ORIGIN

Alignment Scores:
Pred. No.: 1.8e-33 Length: 595
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-09-774-954-9 (1-61) x CN348926 (1-595)

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyAtrg 20
 Db 99 ATGCTTGGGGGCTCTCTGGAGCCCGCGGTACTCTGCTCTACTGCTCCCTGCAATGGGGCGC 158
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnAtrg 40
 Db 159 TTGGGAACAGCGCGATCACTTCTGGCTCTCTGGCATTTGCAAGCTGCTTAACCGT 218
 QY 41 ThrLeuAlaValProProTTPilleGlyTyrGlnHisLysProProPheThrAsnLeu 60
 Db 219 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAGCGCTCTCTTACCACCTC 278

QY 61 His 61

Db 279 CAT 281

RESULT 6
 LOCUS AL555719 610 bp mRNA linear EST 30-MAR-2004
 DEFINITION AL555719 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 CDNA clone CSODK003YN13 5-PRIME, mRNA sequence.

ACCESSION AL555719 GI:45860448

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 610)

Li, W.B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 15, 2001 this sequence version replaced gi:31277525.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

9980.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CSODK003CG07QPl&c=9980.f.

Location/Qualifiers

1..610

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODK003YN13"

/cell_type="HELA CELLS COT 25-NORMALIZED"

/cell_line="HELA"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 1.85e-33 Length: 610

Score: 353.00 Matches: 61

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 1 Gaps: 0

QY 1 MetProAlaGlySerTTPAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyAtrg 20
 Db 113 ATGCTTGGGGGCTCTCTGGAGCCCGCGGTACTCTGCTCTACTGCTCCCTGCAATGGGGCGC 172
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnAtrg 40
 Db 173 TTGGGAACAGCGCGATCACTTCTGGCTCTCTGGCATTTGCAAGCTGCTTAACCGT 232
 QY 41 ThrLeuAlaValProProTTPilleGlyTyrGlnHisLysProProPheThrAsnLeu 60
 Db 233 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAAGCATCAAGCGCTCTCTTACCACCTC 292
 QY 61 His 61
 Db 293 CAT 295

RESULT 7

LOCUS CV029252

DEFINITION

7968 Full length cDNA from the Mammalian Gene Collection Homo

sapiens cDNA 5' similar to BC005082, mRNA sequence.

ACCESSION CV029252

VERSION

CV029252.1 GI:51487402

KEYWORDS

SOURCE

EST.

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 656)

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S.,

Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O.,

Clingingsmith, T.R., Hartley, J.L., Esposito, D., Cheo, D., Moore, T.,

Simmons, B., Sequerra, R., Bosak, S., Doucette-Stamm, L., Le Peuch, C.,

Vandenhaute, J., Cusick, M.E., Albaladejo, J.S., Hill, D.E. and Vidal, M.

Human ORFome Version 1.1: a platform for Reverse Proteomics

Genome Res. (2004) In press

Contact: Vidal M

Marc Vidal Laboratory

Dana Farber Cancer Institute

1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA

Tel: 617 632 5180

Fax: 617 632 5739

Email: Marc.Vidal@dfci.harvard.edu

ORF Sequence Tag (OST) of Gateway Entry construct. Each cloned ORF

results from a PCR reaction using an MGC full-length cDNA as

template DNA and ORF specific primers

PCR Primers

FORWARD: ATGGGCGCGCGCGTGGCG

BACKWARD: CACCTTGGAGAGTAAGTA

Insert Length: 656 Std Error: 18.00

Plate: 11019 row: 04 column: B

Seq primer: ACTGGCCCTCGTTTACACGCTCGTGCCTGGGAAAAAC

High quality sequence start: 87

High quality sequence stop: 655

POLYA=NO.

Location/Qualifiers

1..656

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="mixed"

/clone_lib="Full length cDNA from the Mammalian Gene

Collection"

/note="Vector: mixed; The ORFs were PCR amplified from the

MGC (Mammalian Gene Collection) as of April 2004 and

cloned by recombinational Gateway cloning into pDONR223

Donor vector. Reference: MGC (Mammalian Gene Collection)

Program Team, Generation and Initial Analysis of more than

15,000 Full-Length Human and Mouse cDNA Sequences. FNAS,

2002, 99(26), 16899-16903"

ORIGIN

Alignment Scores:

US-09-774-954-9 (1-61) x AU137536 (1-699)		US-09-774-954-9 (1-61) x CV029252 (1-656)	
Qy	1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20	Qy	1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db	132 ATGCCCGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCCGTCATGGGGCGC 191	Db	70 ATGCTTGGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCCGTCATGGGGCGC 129
Qy	21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40	Qy	21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db	192 TTTGGGAACGAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 251	Db	130 TTTGGGAACGAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 189
Qy	41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60	Qy	41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
Db	252 ACCTTGGCTGCTCCCTCTTGGATTGATACCGATCACAAGCTCTCTTCCACCACTC 311	Db	190 ACCTTGGCTGCTCCCTCTTGGATTGATACCGATCACAAGCTCTCTTCCACCACTC 249
Qy	61 His 61	Qy	61 His 61
Db	312 CAT 314	Db	250 CAT 252
RESULT 9		RESULT 8	
CN348925		AU137536	
LOCUS	705 bp mRNA linear EST 16-MAY-2000	LOCUS	699 bp mRNA linear EST 02-AUG-2002
DEFINITION	17000532814105 GRN_EB Homo sapiens cDNA 5', mRNA sequence.	DEFINITION	AU137536 PLACE1 Homo sapiens cDNA clone PLACE1006691 5', mRNA sequence.
ACCESSION	CN348925	ACCESSION	AU137536
VERSION	CN348925.1 GI:47348859	VERSION	AU137536.1 GI:10998075
KEYWORDS	EST.	KEYWORDS	EST.
SOURCE	Homo sapiens (human)	SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens	ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 705) Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.	AUTHORS	1 (bases 1 to 699) Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Watanabe, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.
TITLE	Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation	TITLE	HRI human cDNA project
JOURNAL	Nat. Biotechnol. 22 (6), 707-716 (2004)	JOURNAL	Unpublished (2000)
COMMENT	Contact: Brandenberger R Regenerative Medicine Geron Corporation 230 Constitution Drive, Menlo Park, CA 94025, USA Tel: 650 473 8658 Fax: 650 473 7760 Email: rbrandenberger@geron.com	COMMENT	Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES	source Location/Qualifiers 1..705 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="embryonic stem cells, embryoid bodies derived from H1, H7 and H9 cells" /clone_lib="GRN_EB" /note="oligo dt-primed, full-length enriched cDNA library from embryoid body outgrowths derived from hES cell lines H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free conditions."	FEATURES	source Location/Qualifiers 1..699 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone_lib="PLACE1006691" /tissue_type="placenta" /clone_lib="PLACE1" /note="Vector: pME18SFL3"
ORIGIN		ORIGIN	
Alignment Scores:		Alignment Scores:	
Pred. No.:	2.21e-33	Pred. No.:	2.18e-33
Score:	353.00	Score:	353.00
Percent Similarity:	100.00%	Percent Similarity:	100.00%
Best Local Similarity:	100.00%	Best Local Similarity:	100.00%
Query Match:	100.00%	Query Match:	100.00%
DB:	7	DB:	1
Length:	656	Length:	699
Matches:	61	Matches:	61
Conservative:	0	Conservative:	0
Mismatches:	0	Mismatches:	0
Indels:	0	Indels:	0
Gaps:	0	Gaps:	0
US-09-774-954-9 (1-61) x CV029252 (1-656)		US-09-774-954-9 (1-61) x CN348925 (1-705)	
Qy	1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20	Qy	1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db	70 ATGCTTGGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCCGTCATGGGGCGC 129	Db	50 ATGCCCGGGCTCTCTGGGACCGCGGCTTACTGCTCTACTGCCCCGTCATGGGGCGC 109
Qy	21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40	Qy	21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db	130 TTTGGGAACGAGCCGATCACTTCTTGGGCTCTCTGGCAATTTGCAAGCTGCTAAACCGT 189		
Qy	41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60		
Db	190 ACCTTGGCTGCTCCCTCTTGGATTGATACCGATCACAAGCTCTCTTCCACCACTC 249		
Qy	61 His 61		
Db	250 CAT 252		
RESULT 8		RESULT 9	
AU137536		CN348925	
LOCUS	699 bp mRNA linear EST 02-AUG-2002	LOCUS	705 bp mRNA linear EST 16-MAY-2000
DEFINITION	AU137536 PLACE1 Homo sapiens cDNA clone PLACE1006691 5', mRNA sequence.	DEFINITION	17000532814105 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION	AU137536	ACCESSION	CN348925
VERSION	AU137536.1 GI:10998075	VERSION	CN348925.1 GI:47348859
KEYWORDS	EST.	KEYWORDS	EST.
SOURCE	Homo sapiens (human)	SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens	ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 699) Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Watanabe, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T.	AUTHORS	1 (bases 1 to 705) Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J., Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R., Lebkowski, J. and Stanton, L. W.
TITLE	HRI human cDNA project	TITLE	Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL	Unpublished (2000)	JOURNAL	Nat. Biotechnol. 22 (6), 707-716 (2004)

```

Db      110 TTTGGGAACCGCGCATCTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGT 169
Qy      41 ThrLeuAlaValProProTfPilleGluTyrglnHisHisLysProProPheThrAsnLeu 60
Db      170 ACCTTGCTGTCCCTCCTTGGATTGAGTACCAAGCATCAAGAGCTCTCTTACCAACCTC 229
Qy      61 His 61
Db      230 CAT 232

RESULT 10
LOCUS   BG424731 753 bp mRNA linear EST 14-MAR-2001
DEFINITION 602453575F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4591867 5',
mRNA sequence.
ACCESSION BG424731
VERSION   BG424731.1 GI:13331237
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 753)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: DCTD/FTP
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov. e column: 20
          Plate: LLC91329 row: e column: 20
          High quality sequence stop: 599.
          Location/Qualifiers
            1..753
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:4591867"
              /tissue_type="renal cell adenocarcinoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_14"
              /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dt priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5'
              adaptor: GGCACGAG(G). Size-selected >500bp for average
              insert size 1.8kb. Library constructed by Ling Hong in
              the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."
```

```

ORIGIN
Alignment Scores:
Pred. No.: 2,39e-33 Length: 753
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-9 (1-61) x BG424731 (1-753)
Qy      1 MetProAlaGlySerTTPAspProAlaGlyTyrluLeuTyrcysProCysMetGlyArg 20
Db      121 ATGCCTGCGGGCTCTCTGGAGCCCGCGGTTACTGCTACTGCCCTGCATGGGGCGC 180
Qy      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      230 CAT 232
```

```

Db      181 TTTGGGAACCGCGCATCTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGT 240
Qy      41 ThrLeuAlaValProProTfPilleGluTyrglnHisHisLysProProPheThrAsnLeu 60
Db      241 ACCTTGCTGTCCCTCCTTGGATTGAGTACCAAGCATCAAGAGCTCTCTTACCAACCTC 300
Qy      61 His 61
Db      301 CAT 303

RESULT 11
LOCUS   BG519639 790 bp mRNA linear EST 02-APR-2001
DEFINITION 602578756F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503461 5',
mRNA sequence.
ACCESSION BG519639
VERSION   BG519639.1 GI:13515377
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 790)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov. e column: 14
          Plate: LLC9176 row: o column: 14
          High quality sequence stop: 710.
          Location/Qualifiers
            1..790
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:3503461"
              /tissue_type="neuroblastoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_19"
              /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dt priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5'
              adaptor: GGCACGAG(G). Library constructed by Ling Hong
              in the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
              (Stratagene) and Superscript II RT (Life Technologies)."
```

```

ORIGIN
Alignment Scores:
Pred. No.: 2,53e-33 Length: 790
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-9 (1-61) x BG519639 (1-790)
Qy      1 MetProAlaGlySerTTPAspProAlaGlyTyrluLeuTyrcysProCysMetGlyArg 20
Db      120 ATGCCTGCGGGCTCTCTGGAGCCCGCGGTTACTGCTACTGCCCTGCATGGGGCGC 179
Qy      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      180 TTTGGGAACCGCGCATCTTCTGGGCTCTGGCATTTGCAAGCTGCTAAACCGT 239
```

QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
 |||||
 Db 240 ACCTTGGCTGCTCCTTGGATTGAGTACCAAGCATCACAGGCTCTTTCACCAACCTC 299
 |||||
 QY 61 His 61
 |||||
 Db 300 CAT 302

RESULT 12
 BG747760 808 bp mRNA linear EST 15-MAY-2001
 LOCUS 602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
 DEFINITION mRNA sequence.
 ACCESSION BG747760
 VERSION BG747760.1 GI:14058413
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 808)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 EMAIL: cgapbs-r@mail.nih.gov
 TISSUE Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1675 row: n column: 14
 High quality sequence stop: 808.

FEATURES
 source
 1..808
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4842061"
 /tissue_type="normal pigmented retinal epithelium"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_43"
 /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library. |"

ORIGIN
 Alignment Scores: 2.6e-33 Length: 808
 Pred. No.: 353.00 Matches: 61
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 100.00% Gaps: 0
 DB: 4

US-09-774-954-9 (1-61) x BG747760 (1-808)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 |||||
 Db 82 ATGCCCGGGCTCTGGACCCGGCGGTACTGCTCTACTGCTGATGGGGCGC 141
 |||||
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 |||||
 Db 142 TTTGGGAACCAAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 201
 |||||
 QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60

Db 202 ACCTTGGCTGCTCCTTGGATTGAGTACCAAGCATCACAGGCTCTTTCACCAACCTC 261
 |||||
 QY 61 His 61
 |||||
 Db 262 CAT 264

RESULT 13
 BX345989 841 bp mRNA linear EST 23-APR-2004
 LOCUS BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
 ACCESSION BX345989
 VERSION BX345989.2 GI:46549776
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 841)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30377055.
 CONTACT Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 9980.f
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?S=CS0AJ008DC06QPl&c=9980.f.

FEATURES
 Location/Qualifiers
 1..841
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ008YF12"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
 10-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
 Alignment Scores: 2.73e-33 Length: 841
 Pred. No.: 353.00 Matches: 61
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 100.00% Indels: 0
 Query Match: 100.00% Gaps: 0
 DB: 5

US-09-774-954-9 (1-61) x BX345989 (1-841)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
 |||||
 Db 109 ATGCTCGGGCTCTGGACCCGGCGGTACTGCTCTACTGCTGATGGGGCGC 168
 |||||
 QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
 |||||
 Db 169 TTTGGGAACCAAGCCGATCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 228
 |||||
 QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
 |||||
 Db 229 ACCTTGGCTGCTCCTTGGATTGAGTACCAAGCATCACAGGCTCTTTCACCAACCTC 288

```

QY      61 His 61
Db      289 CAT 291

RESULT 14
BI524031
LOCUS   603052202F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
DEFINITION mRNA sequence.
ACCESSION BI524031
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 889)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: Life Technologies, Inc.
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM11505 row: d column: 01
  High quality sequence stop: 880.
  Location/Qualifiers
    1..889
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5201616"
    /lab_host="DH10B"
    /clone_lib="NIH_MGC_122"
    /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
    Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
    anonymous pool of 24 week female lung, 16 week female
    spleen, and 20-22 week male spleens. Library is oligo-dT
    primed and directionally cloned (EcoRV site is destroyed
    upon cloning). Average insert size 1.4 kb, insert size
    range 1-3 kb. Library is normalized and enriched for
    full-length clones and was constructed by C. Gruber
    (Invitrogen). Research Genetics tracking code 026. Note:
    this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      2,928-33      Length:      889
Score:          353.00      Matches:    61
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              4          Gaps:        0

US-09-774-954-9 (1-61) x BI524031 (1-889)

QY      1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db      87 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTCTACTGCTGATGGGGCGC 146
QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      147 TTTGGGACACGCGGATCACTCTTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGT 206
QY      41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db      207 ACCTTGGCTGTCCCTCTGGATTGATTGATACGACATCAACGCTCTCTTCCACCACTC 266

```

```

QY      61 His 61
Db      267 CAT 269

RESULT 15
BF315757
LOCUS   601900828F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129975 5',
DEFINITION mRNA sequence.
ACCESSION BF315757
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 898)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: ATCC
  CDNA Library Preparation: Ling Hong/Rubin Laboratory
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
  Plate: LICM1025 row: h column: 08
  High quality sequence stop: 721.
  Location/Qualifiers
    1..898
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:4129975"
    /tissue_type="neuroblastoma"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH_MGC_19"
    /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
    EcoRI; CDNA made by oligo-dT priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5'
    adaptor: GGCACGAG(G). Library constructed by Ling Hong
    in the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      2,96e-33      Length:      898
Score:          353.00      Matches:    61
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              2          Gaps:        0

US-09-774-954-9 (1-61) x BF315757 (1-898)

QY      1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db      120 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTCTACTGCTGATGGGGCGC 179
QY      21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db      180 TTTGGGACACGCGGATCACTCTTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGT 239
QY      41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db      240 ACCTTGGCTGTCCCTCTGGATTGATTGATACGACATCAACGCTCTCTTCCACCACTC 299

```

Db |||
 300 CAT 302

Search completed: October 26, 2005, 15:17:27
Job time : 1069.43 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 ; Search time 260.659 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLFCMGR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9772377 seqs, 4126317084 residues
Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat_25102005_105434_6433/app_query.fasta_1.917
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09774954 @CGN_1_1_684 @runat_25102005_105434_6433
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	1100	11	US-09-774-954-16
2	353	100.0	1300	11	US-09-774-954-4
3	353	100.0	1400	22	US-10-956-157-9986 <
4	353	100.0	1506	22	US-10-956-157-4751 <
5	353	100.0	1514	11	US-09-774-954-1
6	353	100.0	5218	16	US-10-301-822-168 <
7	353	100.0	5218	22	US-10-956-157-4752 <
8	353	100.0	11284	11	US-09-774-954-5
9	223	63.2	1209	26	US-11-097-143-19161
10	218	61.8	5009	11	US-09-774-954-7
11	188	53.3	3264	26	US-11-097-143-18160
12	188	53.3	3793	26	US-11-097-143-18082
13	120	34.0	521	13	US-09-925-065A-502444
14	120	34.0	760	24	US-10-820-474A-236
15	90	25.5	591	9	US-09-864-761-9418
16	90	25.5	591	17	US-10-029-386-24815
17	90	25.5	591	17	US-10-029-386-24815
18	86	24.4	447	21	US-10-425-115-142368
19	82	23.2	402	9	US-09-783-590-11501
20	80.5	22.8	829	19	US-10-424-599-123276
21	77.5	22.0	568	21	US-10-425-115-16666
22	77.5	22.0	1617	20	US-10-437-963-11436
23	75.5	21.4	277616	20	US-10-367-094-83
24	73.5	20.8	1803	16	US-10-156-761-5001
25	73.5	20.8	9025608	16	US-10-156-761-1
26	73	20.7	3545	26	US-11-097-143-11746
27	72	20.4	430	9	US-09-960-352-10634
28	72	20.4	432	9	US-09-960-352-1656
29	72	20.4	447	9	US-09-960-352-6689
30	71	20.1	1186	18	US-10-282-122A-26446
31	71	20.1	1243	18	US-10-282-122A-26446
32	71	20.1	4953	20	US-10-437-963-23967
33	70.5	20.0	467	10	US-09-918-995-29055
34	70.5	20.0	510	17	US-10-029-386-10060
35	70	19.8	326002	22	US-10-461-862-56
36	69.5	19.7	622	20	US-10-437-963-15591
37	69	19.5	600	24	US-10-972-079-32895
38	68.5	19.4	600	24	US-10-972-079-43878
39	68.5	19.4	600	24	US-10-972-079-43879
40	68.5	19.4	600	24	US-10-972-079-43880
41	68.5	19.4	3373	15	US-10-105-963-9
42	68.5	19.4	8174	9	US-09-863-475A-5
43	67.5	19.1	358	21	US-10-425-115-50448
44	67.5	19.1	600	24	US-10-972-079-27372
45	67.5	19.1	636	13	US-09-925-065A-285729

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-Nov-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-Jan-1997
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1100 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

Alignment Scores:
Pred. No.: 5,828-40 Length: 1100
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-9 (1-61) x US-09-774-954-16 (1-1100)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 1 ATGCCCGCGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCCTGCATGGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 61 TTTGGGAACCCAGCCGATCACCTCTTGGGCTCTGGCATTTTGCAAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProProTrpIleGlyTyrGlnHisLysProProPheThrAsnLeu 60
DB 121 ACCTTGGCTCTCCCTCTGGATTGAGTACCAAGCATCACAAAGCCTCTTTCACCAACCTC 180
QY 61 His 61
DB 181 CAT 183

RESULT 2
US-09-774-954-4
Sequence 4, Application US/09774954
Publication No. US20040241645A1
GENERAL INFORMATION:
APPLICANT: Yang Wang, Michael W. Spellman
TITLE OF INVENTION: O-Fucosyltransferase
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-Nov-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-Jan-1997
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1300 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

Alignment Scores:
Pred. No.: 7,1e-40 Length: 1300
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-9 (1-61) x US-09-774-954-4 (1-1300)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
DB 136 ATGCCCGCGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCCTGCATGGGGCGC 195
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
DB 196 TTTGGGAACCCAGCCGATCACCTCTTGGGCTCTGGCATTTTGCAAAAGCTGCTAAACCGT 255
QY 41 ThrLeuAlaValProProTrpIleGlyTyrGlnHisLysProProPheThrAsnLeu 60
DB 256 ACCTTGGCTCTCCCTCTGGATTGAGTACCAAGCATCACAAAGCCTCTTTCACCAACCTC 315
QY 61 His 61
DB 316 CAT 318

RESULT 3
US-10-956-157-9986
Sequence 9986, Application US/10956157
Publication No. US20050118625A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
FILE REFERENCE: 031896-043000 (AM 101081)
CURRENT APPLICATION NUMBER: US/10/956,157
NUMBER OF SEQ ID NOS: 2004-10-04
CURRENT FILING DATE: 2004-10-04
NUMBER OF SEQ ID NOS: 319805
SOFTWARE: PatentIn version 3.2
SEQ ID NO 9986
LENGTH: 1400
TYPE: DNA
ORGANISM: Homo sapiens

US-10-956-157-9986

Alignment Scores:
Pred. No.: 7.75e-40 Length: 1400
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-9 (1-61) x US-10-956-157-9986 (1-1400)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuTyrCysProCysMetGlyArg 20
Db 13 ATGCTGCGGGCTCTTGGACCGCGGTACTGCTCTACTGCGCGCTGATGGGGCGC 72
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerIeuAlaPheAlaLysLeuLeuAsnArg 40
Db 73 TTGGGAACACGACCGCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 132
QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisIleHisLysProProPheThrAsnLeu 60
Db 133 ACCTTGGCTGCTCCCTCTGGATTGAGTACGAGCATCACAAAGCTCTCTTCAACCACTC 192
QY 61 His 61
Db 193 CAT 195

RESULT 4

US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:
Pred. No.: 8.45e-40 Length: 1506
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-9 (1-61) x US-10-956-157-4751 (1-1506)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuTyrCysProCysMetGlyArg 20
Db 119 ATGCTGCGGGCTCTTGGACCGCGGTACTGCTCTACTGCGCGCTGATGGGGCGC 178
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerIeuAlaPheAlaLysLeuLeuAsnArg 40
Db 179 TTGGGAACACGACCGCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 238
QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisIleHisLysProProPheThrAsnLeu 60
Db 239 ACCTTGGCTGCTCCCTCTGGATTGAGTACGAGCATCACAAAGCTCTCTTCAACCACTC 298
QY 61 His 61
Db 299 CAT 301

RESULT 5

US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Alignment Scores:
Pred. No.: 8.51e-40 Length: 1514
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-09-774-954-9 (1-61) x US-09-774-954-1 (1-1514)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCTGCGGGCTCTTGGACCGCGGTACTGCTCTACTGCGCGCTGATGGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerIeuAlaPheAlaLysLeuLeuAsnArg 40
Db 61 TTGGGAACACGACCGCATCTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProProTrpIleGluTyrGlnHisIleHisLysProProPheThrAsnLeu 60
Db 121 ACCTTGGCTGCTCCCTCTGGATTGAGTACGAGCATCACAAAGCTCTCTTCAACCACTC 180
QY 61 His 61
Db 181 CAT 183
RESULT 6

```
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029F2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 5218
; TYPE: DNA
; ORGANISM: Homo Sapiens
; NAME/KEY: CDS
; LOCATION: (50)...(1216)
US-10-301-822-168

Alignment Scores:
Pred. No.: 3,7e-39 Length: 5218
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

US-09-774-954-9 (1-61) x US-10-301-822-168 (1-5218)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 119 ATGCCTGCGGGCTCTCTGGGACCCGCGGTACTCTGCTCTACTGCCCCGATGGGGCGC 178
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 179 TTTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 238
QY 41 ThrLeuAlaValProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
Db 239 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCACAAAGCCTCTTTTACCAACCTC 298
QY 61 His 61
Db 299 CAT 301

RESULT 7
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
```

```
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores:
Pred. No.: 3,7e-39 Length: 5218
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-774-954-9 (1-61) x US-10-956-157-4752 (1-5218)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 119 ATGCCTGCGGGCTCTCTGGGACCCGCGGTACTCTGCTCTACTGCCCCGATGGGGCGC 178
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 179 TTTGGGAACACGAGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGT 238
QY 41 ThrLeuAlaValProTrpIleGluTyrGlnHisHisLysProPheThrAsnLeu 60
Db 239 ACCTTGGCTGTCCCTCTTGGATTGAGTACCAGCATCACAAAGCCTCTTTTACCAACCTC 298
QY 61 His 61
Db 299 CAT 301

RESULT 8
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11284 base pairs
```

Alignment Scores:		
pred. No.:	2.25e-21	Length: 1209
Score:	223.00	Matches: 38
Percent similarity:	91.11%	Conservative: 3

```
QY 23 AspGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
Db 2 AACAGGCGGATCACTCTTGGGCTCTCGCAATTCGAAAGCTGCTAAACCGTACTTIG 61
QY 43 AlaValProProTTPileGluTyrGlnHisHisGlyProProPheThrAsnLeuHis 61
Db 62 GCTGTCCCTCCCTTGGATTGAGTACGAGCATCAAGGCTCTTTACCAACCTCCAT 118

RESULT 11
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 18160
; LENGTH: 3264
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.: 7 178-16 Length: 3264
Score: 188.00 Matches: 36
Percent Similarity: 64.06% Conservative: 3
Best Local Similarity: 59.38% Mismatches: 4
Query Match: 53.26% Indels: 19
DB: 26 Gaps: 1

US-09-774-954-9 (1-61) x US-11-097-143-18160 (1-3264)
QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
Db 2183 GATCCCAATGGCTACCTCACCTACTGTCGCTGATGGGTAAGTTTCACTTGAAGCAACC 2124
QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db 2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGAGCGCTTTGGCAACCAAGCGCCGACCACTT 2064
QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTTPil 48
Db 2063 CTGGGATCATTTGGCTTTGGCAAGGCGCTTAATCGCACCCCTGATCCTGCGCGCGTGGGT 2004
QY 48 eGluTyrGln 51
Db 2003 GGAGTATCGT 1994

RESULT 12
US-11-097-143-18082/c
; Sequence 18082, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 18082
; LENGTH: 3793
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.: 8 578-16 Length: 3793
Score: 188.00 Matches: 38
Percent Similarity: 64.06% Conservative: 3
Best Local Similarity: 59.38% Mismatches: 4
Query Match: 53.26% Indels: 19
DB: 26 Gaps: 1

US-09-774-954-9 (1-61) x US-11-097-143-18082 (1-3793)
QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
Db 675 GATCCCAATGGCTACCTCACCTACTGTCGCTGATGGGTAAGTTTCACTTGAAGCAACC 616
QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db 615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGAGCGCTTTGGCAACCAAGCGCCGACCACTT 556
QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTTPil 48
Db 555 CTGGGATCATTTGGCTTTGGCAAGGCGCTTAATCGCACCCCTGATCCTGCGCGCGTGGGT 496
QY 48 eGluTyrGln 51
Db 495 GGAGTATCGT 486

RESULT 13
US-09-925-065A-602444/c
; Sequence 602444, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
```

;; PRIOR FILING DATE: 2000-10-24
;; PRIOR APPLICATION NUMBER: US 60/252,147
;; PRIOR FILING DATE: 2000-11-20
;; PRIOR APPLICATION NUMBER: US 60/250,092
;; PRIOR FILING DATE: 2000-11-30
;; PRIOR APPLICATION NUMBER: US 60/261,766
;; PRIOR FILING DATE: 2001-01-16
;; PRIOR APPLICATION NUMBER: US 60/289,846
;; PRIOR FILING DATE: 2001-05-09
;; NUMBER OF SEQ ID NOS: 957086
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 602444
;; LENGTH: 521
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-925-065A-602444

Alignment Scores:
Pred. No.: 4,03e-07 Length: 521
Score: 120.00 Matches: 19
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 33.99% Indels: 0
DB: 13 Gaps: 0

US-09-774-954-9 (1-61) x US-09-925-065A-602444 (1-521)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 286 ATGCCVGGGGCTCTCTGGACCGCGGTACCTGCTTACTGCGCCCTGCATGGGTAAG 227

RESULT 14
US-10-820-474A-236
; Sequence 236, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
; APPLICANT: LAL, PREETI
; APPLICANT: TANG, Y. TOM
; APPLICANT: GORGONE, GINA A.
; APPLICANT: CORLEY, NEIL C.
; APPLICANT: GUEGLER, KARL J.
; APPLICANT: BAUGHN, MARIAH R.
; APPLICANT: AKERBLOM, INGRID E.
; APPLICANT: AU-YOUNG, JANICE
; APPLICANT: YUE, HENRY
; APPLICANT: REDDY, ROOPA
; APPLICANT: PATTERSON, CHANDRA
; APPLICANT: HILLMAN, JENNIFER L.
; APPLICANT: BANDMAN, OLGA
; TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
; FILE REFERENCE: 039386-1568
; CURRENT APPLICATION NUMBER: US/10/820,474A
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: 09/720,533
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: PCT/US99/14484
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/090,762
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/094,983
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: 60/102,686
; PRIOR FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
; LENGTH: 760
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone No: 1647884

Alignment Scores:
Pred. No.: 6.32e-07 Length: 760
Score: 120.00 Matches: 19
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 33.99% Indels: 0
DB: 24 Gaps: 0

US-09-774-954-9 (1-61) x US-10-820-474A-236 (1-760)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 111 ATGCCCGGGCTCTCTGGACCGCGGTACCTGCTTACTGCCCCCTGCATGGGTAAG 170

RESULT 15
US-09-864-761-9418
; Sequence 9418, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9418
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: MAP TO AC004624.6

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
US-09-864-761-9418

Alignment Scores:
Pred. No.: 0.00889 Length: 591
Score: 90.00 Matches: 21
Percent Similarity: 41.18% Conservative: 7
Best Local Similarity: 30.88% Mismatches: 18
Query Match: 25.50% Indels: 22
Db: 9 Gaps: 2

US-09-774-954-9 (1-61) x US-09-864-761-9418 (1-591)

QY 2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
Db 387 CCGAGGGGAGTGGGACCCCGGGGCTGCTCAGACTACAGGGTCTGCCAGCGCCA 446
QY 21 PheGlyAsnGlnAlaAsp----- 26
Db 447 TTTGGGCACTCCAGTGACCCAGGGCTCGGGTCTCGCACACCTCCCATGAGGAGATGTC 506
QY 27 -----HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
Db 507 TTCTCATGGAATCAAAGCCACATCTGTGGTTCTATCTCTGGGACCAGGTCTGCAGA 566
QY 40 ArgThrLeuAlaValProProTip 47
Db 567 GCTGGTTTAGGTCATCCTCTTGG 590

Search completed: October 26, 2005, 16:00:15
Job time : 272.326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: October 26, 2005, 08:18:39 ; Search time 53.4846 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09774954/runat_25102005_105433_6415/app_query.fasta_1.917
-DB=Issued Patents NA -CFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954 @CGN 1.1.16 @runat_25102005_105433_6415 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	1100	3	US-08-978-741-16
2	353	100.0	1100	3	US-09-333-729A-16
3	353	100.0	1300	3	US-08-978-741-4
4	353	100.0	1300	3	US-09-333-729A-6
5	353	100.0	1514	3	US-08-978-741-1
6	353	100.0	1514	3	US-09-333-729A-2
7	353	100.0	11284	3	US-08-978-741-5
8	223	63.2	1320	4	US-09-270-767-14353
9	218	61.8	5009	3	US-08-978-741-7
10	218	61.8	5009	3	US-09-333-729A-8
11	120	34.0	19227	4	US-09-949-016-12127
12	120	34.0	19228	4	US-09-949-016-16285

c	13	75.5	21.4	260247	4	US-09-949-016-13358	Sequence 13358, A
	14	71	20.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	15	71	20.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	16	70.5	20.0	36759	4	US-09-949-016-12216	Sequence 12216, A
	17	70.5	20.0	36760	4	US-09-949-016-14021	Sequence 14021, A
	18	68.5	19.4	999	5	PCT-US91-00899-10	Sequence 10, Appl
	19	68.5	19.4	1136	1	US-08-395-800A-9	Sequence 9, Appli
	20	68.5	19.4	1155	1	US-08-434-151-1	Sequence 1, Appli
	21	68.5	19.4	1155	1	US-08-208-889A-1	Sequence 1, Appli
	22	68.5	19.4	1155	2	US-08-433-271-1	Sequence 1, Appli
	23	68.5	19.4	1155	2	US-08-715-259-1	Sequence 1, Appli
	24	68.5	19.4	1174	5	PCT-US95-07554-3	Sequence 3, Appli
	25	68.5	19.4	1199	1	US-08-395-800A-5	Sequence 5, Appli
	26	68.5	19.4	2268	1	US-08-675-773B-4	Sequence 4, Appli
	27	68.5	19.4	3373	3	US-08-273-411-2	Sequence 2, Appli
	28	68.5	19.4	3374	3	US-09-949-016-5482	Sequence 5482, Ap
	29	68.5	19.4	3791	3	US-08-675-773B-3	Sequence 3, Appli
	30	68.5	19.4	8174	1	US-07-914-281-5	Sequence 5, Appli
	31	68.5	19.4	8174	1	US-08-393-246-5	Sequence 5, Appli
	32	68.5	19.4	8174	1	US-08-525-058A-5	Sequence 5, Appli
	33	68.5	19.4	8174	3	US-08-696-731-5	Sequence 5, Appli
	34	68.5	19.4	8174	3	US-09-042-531-5	Sequence 5, Appli
	35	68.5	19.4	8174	5	PCT-US91-00899-3	Sequence 3, Appli
	36	68.5	19.4	8580	4	US-09-949-016-17224	Sequence 17224, A
	37	66.5	18.8	26896	4	US-09-949-016-16800	Sequence 16800, A
	38	66	18.7	1029	4	US-09-244-805-60	Sequence 60, Appl
	39	65.5	18.6	87644	4	US-09-949-016-16041	Sequence 16041, A
	40	65	18.4	601	4	US-09-949-016-84839	Sequence 84839, A
	41	65	18.4	321022	4	US-09-949-016-84840	Sequence 84840, A
	42	65	18.4	321022	4	US-09-949-016-11852	Sequence 11852, A
	43	65	18.4	321022	4	US-09-949-016-14166	Sequence 14166, A
	44	64.5	18.3	40429	4	US-08-311-731A-125	Sequence 125, App
	45	64.5	18.3	421491	4	US-09-949-016-12805	Sequence 12805, A

ALIGNMENTS

RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 1100 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-16
Alignment Scores:
Pred. No.: 1198-39 Length: 1100
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-08-978-741-16 (1-1100)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTACTGCGCCCTGCATGGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 61 TTGGGAACACCGCGCATCACTTCTGGGCTCTCTGGCATTTGCCAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 180
QY 61 His 61
Db 181 CAT 183

RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCES: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 16
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-16
Alignment Scores:
Pred. No.: 1198-39 Length: 1100
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-09-333-729A-16 (1-1100)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCCCGCGGGCTCTGGGACCCGCGGTACCTGCTACTGCGCCCTGCATGGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 61 TTGGGAACACCGCGCATCACTTCTGGGCTCTCTGGCATTTGCCAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 180
QY 61 His 61
Db 181 CAT 183
```

```
QY 61 His 61
Db 181 CAT 183

RESULT 3
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 1300 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-4
Alignment Scores:
Pred. No.: 15e-39 Length: 1300
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-08-978-741-4 (1-1300)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 136 ATGCCCGGGCTCTGGGACCCGCGGTACTGCTCTACTGCCCCCTGCATGGGGCGC 195
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 196 TTGGGAACACCGCGCATCACTTCTGGGCTCTCTGGCATTTGCCAAAGCTGCTAAACCGT 255
QY 41 ThrLeuAlaValProProTrrPileGlyTyrGlnHisLysProProPheThrAsnLeu 60
Db 256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAAGCATCAAGGCTCTCTTCCACCAACCTC 315
QY 61 His 61
Db 316 CAT 318
```

```
RESULT 4
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 6
; LENGTH: 1300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6
Alignment Scores:
Pred. No.: 1.5e-39 Length: 1300
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-774-954-9 (1-61) x US-09-333-729A-6 (1-1300)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 136 ATGCCCGCGGGCTCTGGGACCGCGGTACTGCTCTACTGCCCTGCATGGGGCGC 195
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 196 TTGTGGGAACAGGCGGATCACTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 255
QY 41 ThrLeuAlaValProTrpTlleGluTyrGlnHisLysProPheThrAsnLeu 60
Db 256 ACCTTGGCTGCTCTCTTGGATTGAGTACCAGCATCACAAAGCTCCTTTTCCACCAACCTC 315
QY 61 His 61
Db 316 CAT 318
RESULT 5
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-978-741-1
Alignment Scores:
Pred. No.: 1.84e-39 Length: 1514
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-774-954-9 (1-61) x US-08-978-741-1 (1-1514)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCCCGCGGGCTCTGGGACCGCGGTACTGCTCTACTGCCCTGCATGGGGCGC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 61 TTGTGGGAACAGGCGGATCACTCTTGGGCTCTTGGCATTTGCAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProTrpTlleGluTyrGlnHisLysProPheThrAsnLeu 60
Db 121 ACCTTGGCTGCTCTCTTGGATTGAGTACCAGCATCACAAAGCTCCTTTTCCACCAACCTC 180
QY 61 His 61
Db 181 CAT 183
RESULT 6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1D1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo Sapien
US-09-333-729A-2
Alignment Scores:
Pred. No.: 1.84e-39 Length: 1514
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-774-954-9 (1-61) x US-09-333-729A-2 (1-1514)
```

```
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 1 ATGCCCGGGCTCTGGACCCGCGGTACTGCTCTACTGCCCTGCATGGGGCC 60
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db 61 TTTGGGAACCAAGCGGATCACTTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 120
QY 41 ThrLeuAlaValProTrpTyrPheGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db 121 ACCTTGGCTGTCCCTCTCTGGATTGAGTACCAAGCATCAAGGCTCTCTTTACCAACCTC 180
QY 61 His 61
Db 181 CAT 183

RESULT 7
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11284 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
;
US-08-978-741-5
Alignment Scores:
Pred. No.: 2,818-38 Length: 11284
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-08-978-741-5 (1-11284)
QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 4236 ATGCCCGGGCTCTGGACCCGCGGTACTGCTCTACTGCCCTGCATGGGGCC 4295
QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
```

```
Db 4296 TTTGGGAACCAAGCGGATCACTTTTGGGCTCTCTGGCATTTGCAAAAGCTGCTAAACCGT 4355
QY 41 ThrLeuAlaValProTrpTyrPheGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db 4356 ACCTTGGCTGTCCCTCTCTGGATTGAGTACCAAGCATCAAGGCTCTCTTTACCAACCTC 4415
QY 61 His 61
Db 4416 CAT 4418

RESULT 8
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14353
Alignment Scores:
Pred. No.: 2,188-21 Length: 1320
Score: 223.00 Matches: 38
Percent Similarity: 91.11% Conservative: 3
Best Local Similarity: 84.44% Mismatches: 4
Query Match: 63.17% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-9 (1-61) x US-09-270-767-14353 (1-1320)
QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 25
Db 75 GATCCCAATGGCTACCTCACCTACTGTCGGTATGGGACGCTTTGGCAACACGCGGCAC 134
QY 27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
Db 135 CACTTCTGGGATCAATTGGCTTCCGCAAGGCGCTTAATCCACCTGATCTGCGCGCG 194
QY 47 TrpIleGluTyrGln 51
Db 195 TGGGTGGAGTATCGT 209

RESULT 9
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
```

FILING DATE: 26-NO. 6100076-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA: 08/792498
APPLICATION NUMBER: 08/792498
FILING DATE: 31
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 5009 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
US-08-978-741-7

Alignment Scores:
Pred. No.: 6,64e-20 Length: 5009
Score: 218.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.76% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-08-978-741-7 (1-5009)

Qy 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
Db 2 AACAGGCCGATCATCTTCTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGTACCTTG 61
Qy 43 AlaValProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
Db 62 GCTGTCCCTCTTGGATTGATGACCATCAAGGCTCTCTTTCACCAACCTCCAT 118

RESULT 10

US-09-333-729A-8

Sequence 8, Application US/09333729A

Patent No. 6270987

GENERAL INFORMATION:

APPLICANT: Wang, Yang

APPLICANT: Spellman, Michael W.

TITLE OF INVENTION: O-Fucosyltransferase

FILE REFERENCE: P1041P1D1-Substitute

CURRENT APPLICATION NUMBER: US/09/333,729A

CURRENT FILING DATE: 1999-06-15

PRIOR APPLICATION NUMBER: US 08/798,741

PRIOR FILING DATE: 1997-11-26

NUMBER OF SEQ ID NOS: 21

SEQ ID NO 8

LENGTH: 5009

TYPE: DNA

ORGANISM: Homo Sapien

US-09-333-729A-8

Alignment Scores:

Pred. No.: 6,64e-20 Length: 5009
Score: 218.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.76% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x US-09-333-729A-8 (1-5009)

Qy 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42

Db 2 AACAGGCCGATCATCTTCTGGGCTCTGCGCATTTGCAAGCTGCTAAACCGTACCTTG 61

Qy 43 AlaValProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61

Db 62 GCTGTCCCTCTTGGATTGATGACCATCAAGGCTCTTTCACCAACCTCCAT 118

RESULT 11

US-09-949-016-12127/c

Sequence 12127, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12127

LENGTH: 19227

TYPE: DNA

ORGANISM: Human

US-09-949-016-12127

Alignment Scores:

Pred. No.: 1,99e-05 Length: 19227
Score: 120.00 Matches: 19
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 33.99% Indels: 0
DB: 4 Gaps: 0

US-09-774-954-9 (1-61) x US-09-949-016-12127 (1-19227)

Qy 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20

Db 1718 ATGCCCGGGCTCTTGGGACCGCGGCTTACTGCTCTACTGCTCCCTGCGTGGTAA 1659

RESULT 12

US-09-949-016-16285/c

Sequence 16285, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 16285

LENGTH: 19228

TYPE: DNA

ORGANISM: Human

US-09-949-016-16285

Alignment Scores:

Pred. No.: 1,99e-05 Length: 19228
Score: 120.00 Matches: 19
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 33.99% Indels: 0

```
DB:                                     4          Gaps:                                0
US-09-774-954-9 (1-61) x US-09-949-016-16285 (1-19228)

QY      1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db      1718 ATGCCCGGGGCTCTGGACCGCGGGTACCTGCTCTACTGCCCTGCATGGGTAAAG 1659

RESULT 13
US-09-949-016-13358/c
; Sequence 13358, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13358
; LENGTH: 260247
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13358

Alignment Scores:
Pred. No.:      1,12e+03      Length:      260247
Score:          75.50        Matches:      19
Percent Similarity: 70.27%   Conservative: 7
Best Local Similarity: 51.35% Mismatches:      8
Query Match:    21.39%      Indels:        3
DB:             4           Gaps:          2

US-09-774-954-9 (1-61) x US-09-949-016-13358 (1-260247)

QY      8 ProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAspHis 27
Db      138282 CCCTCCCAATTCCTACTGTTTGTGCC-----AGCGCGCTAGGAATGAGACCAATAT 138229

QY      28 PheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
Db      138228 TTTTATGAAGCCGCTTTTGCT---CTTCTTAAACAACAATTCAGTT 138181

RESULT 14
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: FRASER, Owen R.
; APPLICANT: WHITE, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
```

```
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:      2,1e+05      Length:      4403765
Score:          71.00        Matches:      18
Percent Similarity: 47.62%   Conservative: 2
Best Local Similarity: 42.86% Mismatches:     10
Query Match:    20.11%      Indels:        2
DB:             3           Gaps:          2

US-09-774-954-9 (1-61) x US-09-103-840A-2 (1-4403765)

QY      5 SerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGln 24
Db      222566 AGTTGGACCTTGCT-----GTCTGTATGCCCTTGTCGCGCGGTGACAAACGGTTCC 222619

QY      25 AlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
Db      222620 GCTGGT-----GCGTTGACCGCACACTGGCCGCG 222649

QY      45 ProPro 46
Db      222650 CCGCCG 222655

RESULT 15
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:      2,1e+05      Length:      4411529
Score:          71.00        Matches:      18
Percent Similarity: 47.62%   Conservative: 2
Best Local Similarity: 42.86% Mismatches:     10
Query Match:    20.11%      Indels:        2
DB:             3           Gaps:          2

US-09-774-954-9 (1-61) x US-09-103-840A-1 (1-4411529)

QY      5 SerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGln 24
Db      222457 AGTTGGACCTTGCT-----GTCTGTATGCCCTTGTCGCGCGGTGACAAACGGTTCC 222510

QY      25 AlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
Db      222511 GCTGGT-----GCGTTGACCGCACACTGGCCGCG 222540

QY      45 ProPro 46
Db      222541 CCGCCG 222546

Search completed: October 26, 2005, 15:52:24
```

Job time : 1666.48 secs

THIS PAGE BLANK (ASTO)


```

Db          312 CAT 314
||||
RESULT 4
AAV65633
ID AAV65633 standard; DNA; 11284 BP.
XX
XX
AC AAV65633;
XX
DT 16-DEC-1998 (first entry)
XX
XX Plasmid construct for expression of human O-fucosyltransferase.
XX
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
XX CDS 4140..5333
XX /*tag= a
XX /*note= "insert coding for human O-fucosyltransferase."
XX FT FT 4218..4235
XX FT FT /*tag= b
XX FT FT /*note= "polyhistidine tag"
XX
XX WO9833924-A1.
XX
XX 06-AUG-1998.
XX
XX 17-DEC-1997; 97WO-US023401.
XX
XX 31-JAN-1997; 97US-00792498.
XX 26-NOV-1997; 97US-00978741.
XX
XX (GETH ) GENENTECH INC.
XX
XX Wang Y, Spellman MW;
XX
XX WPI; 1998-437477/37.
XX P-PSDB; AAW80573.
XX
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
XX domains - useful for diagnosis and treatment of diseases involving
XX overexpression of the enzyme.
XX
XX Example; Page 49-57; 90pp; English.
XX
XX This represents the nucleotide sequence of the plasmid construct used for
XX the expression of human heart O-fucosyltransferase. The human O-
XX fucosyltransferase can glycosylate an epidermal growth factor (EGF)
XX domain of a polypeptide with an activated O-fucose residue. Inhibitors of
XX O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
XX domains, are used in diagnosis and treatment of conditions associated
XX with overexpression of O-fucosyltransferase, to promote survival of
XX sensory (retinal) neurons. Probes based on EGF domain polypeptide are
XX used to detect gene amplification and expression. The expression can also
XX be determined at the protein level using antibodies specific for O-
XX fucosyltransferase
XX
XX Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;
SQ

```

Alignment Scores: 5.95e-36 Length: 11284

Pred. No.: 353.00 Matches: 61

Score: 100.00% Conservative: 0

Percent Similarity: 100.00% Mismatches: 0

Best Local Similarity: 100.00% Indels: 0

Query Match: 100.00% Gaps: 0

DB: 2

US-09-774-954-9 (1-61) x AAV65633 (1-11284)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 4236 ATGCCCGCGGGCTCTGGGACCGCGCGGTACTCTGCTCTACTGCGCTGCATGGGGGC 4295

QY 21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLeuLeuAsnArg 40
Db 4296 TTTGGGAACCGCGCGATCACTTCTTGGGCTCTCTGGCATTTGCAAGCTGTAAACCGT 4355

QY 41 ThrLeuAlaValProProTrrPileGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db 4356 ACCTTGGCTGTCCTCTCTGGATTGAGTACCGAGCATCAAGCGCTCTTTTCCACCACTTC 4415

QY 61 His 61
Db 4416 CAT 4418

RESULT 5
ID AAK51510 standard; cDNA; 5230 BP.
XX AAK51510;
XX AAK51510;
DT 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 55.
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00654936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX P-PSDB; AAM78377.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 1; Page 638-642; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111

CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Alignment Scores: 1-87e-21 Length: 5230
Pred. No.: 241.50 Matches: 46
Score: 81.67% Conservative: 3
Percent Similarity: 76.67% Mismatches: 6
Best Local Similarity: 68.41% Indels: 5
Query Match: 4 Gaps: 1
DB:

US-09-774-954-9 (1-61) x AAK51510 (1-5230)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet-----GlyArgPhe 21
AAK52494
ID AAK52494 standard; cDNA; 4850 BP.
XX
XX AAK52494;
XX
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 2023.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00654936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX P-PSDB; AAM79361.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 1; Page 4424-4425; 6221pp; English.

XX This represents a first EcorI nucleotide fragment of human KIA0180. This
CC 5009 basepairs partial cDNA encodes for a protein of unknown function
CC from myeloblast cell line KG-1. The invention provides a human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase

SQ Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,47e-18 Length: 5009
Score: 218.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 61.76% Indels: 0
DB: 2 Gaps: 0

US-09-774-954-9 (1-61) x AAV65634 (1-5009)

QY 23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
DB 2 AACGAGCGCGATCACTTCTGGGCTCTCTGGCATTTGCAAGCTGCTAAACCGTACCTTG 61
QY 43 AlaValProProTyrPileGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
DB 62 GCTGTCCCTCCTTGATTTAGTACCAGATCAGACGCTCTTTTACCAACCTCCAT 118

RESULT 9

ABL13946/c
ID ABL13946 standard; cDNA; 3264 BP.

XX ABL13946;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB69843.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.48e-14 Length: 3264
Score: 188.00 Matches: 38
Percent Similarity: 64.06% Conservative: 3
Best Local Similarity: 59.38% Mismatches: 4
Query Match: 53.26% Indels: 19
DB: 4 Gaps: 1

US-09-774-954-9 (1-61) x ABL13946 (1-3264)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
DB 2183 GATCCCAATGGCTACTCCTCCTACTCTGTCCTGTAAGTTTCACTTGGAGCAACC 2124

QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
DB 2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCGCGCCACTT 2064

QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpI 48
DB 2063 CTTGGGATCATTTGGCTTCGCAAGCGCTTAATCGACCTGATCTCTGCGCGCTGGGT 2004

QY 48 eGluTyrGln 51
DB 2003 GGAGTATCGT 1994

RESULT 10

ABL13894/c
ID ABL13894 standard; cDNA; 3793 BP.

XX ABL13894;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36164.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB69791.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 36164; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3793 BP; 1020 A; 853 C; 950 G; 970 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.79e-14 Length: 3793
Score: 188.00 Matches: 38
Percent Similarity: 64.06% Conservatives: 3
Best Local Similarity: 59.38% Mismatches: 4
Query Match: 53.26% Indels: 19
DB: 4 Gaps: 1

US-09-774-954-9 (1-61) x ABL13894 (1-3793)

QY 7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----- 18
Db 675 GATCCCAATGGCTACCTCCTACTGCTGATGTTAAGTTTCATCTGGAAGCAACC 616
QY 19 -----GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db 615 GAAACACGACACTAAACCGCGCTTTTCAAGAGCGCTTTGGCAACCGAGCCGACCACTT 556
QY 28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpI1 48
Db 555 CTTGGGATCATTTGGCTTCGCCAAGCGCTTAATCGCACCTGATCCTGCCCGCGTGGGT 496
QY 48 eGluTyrGln 51
Db 495 GGAGTATCGT 486

RESULT 11
AAZ98210
ID AAZ98210 standard; cDNA; 760 BP.
XX
AC AAZ98210;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-102 cDNA SEQ ID NO:236.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
FN WO200000610-A2.
XX
XX 06-JAN-2000. 99WO-US014484.
XX
PF 25-JUN-1999;
XX
PR 26-JUN-1998; 98US-0090762P.
PR 31-JUL-1998; 98US-0094983P.
PR 01-OCT-1998; 98US-0102686P.
PR 11-DEC-1998; 98US-0112129P.

PA (INCY-) INCYTE PHARM INC.
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX WPI; 2000-160673/14.
DR P-PSDB; AAY87325.
XX New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT disease.
XX
XX Claim 9; Page 307; 327pp; English.
XX
XX AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC or ribozyme therapeutics, for detecting related sequences or genetic
CC variations, and for chromosomal mapping. HSPP are also used to raise
CC specific antibodies (Ab) and to screen for agonists and antagonists
CC (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC competitive drug screens, and for purification of HSPP from natural
CC sources
XX
SQ Sequence 760 BP; 171 A; 206 C; 220 G; 163 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.9e-06 Length: 760
Score: 120.00 Matches: 19
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 95.00% Mismatches: 0
Query Match: 33.99% Indels: 0
DB: 3 Gaps: 0

US-09-774-954-9 (1-61) x AAZ98210 (1-760)

QY 1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db 111 ATGCCCGCGGCTCTGGGACCGCGCGGTACTCTCTACTGCCCTGTCGCTGAAG 170

RESULT 12
ADQ97980
ID ADQ97980 standard; DNA; 97081 BP.
XX
AC ADQ97980;
XX
XX 07-OCT-2004 (first entry)
XX
XX Human cancer associated sequence HD11-034, SEQ ID 957.
XX
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
XX Homo sapiens.
XX
XX WO2004060304-A2.
XX
XX 22-JUL-2004.
PD

```
XX 22-DEC-2003; 2003WO-US041389.
XX 27-DEC-2002; 2002US-00330773.
XX (SAGR-) SAGRES DISCOVERY INC.
XX Morris DW, Malandro MS;
XX WPI; 2004-543781/52.
XX New isolated cancer associated nucleic acids comprising at least 10
XX contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX cancers such as leukemia and lymphoma.
XX Claim 1; SEQ ID NO 957; 199pp; English.
XX The present invention relates to cancer associated sequences (ADQ97025-
XX ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 97081 BP; 23547 A; 23101 C; 24653 G; 25780 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.33 Length: 97081
Score: 94.00 Matches: 21
Percent Similarity: 43.84% Conservative: 11
Best Local Similarity: 28.77% Mismatches: 19
Query Match: 26.63% Indels: 22
DB: 12 Gaps: 2

US-09-774-954-9 (1-61) x ADQ97980 (1-97081)
QY 2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
DQ 66320 CCGAGGGCAGTTGGAGCCCGCGGGCTCTCAGACTACAGGGTGTGCGCCAGCGCCA 66379
QY 21 PheGlyAsnGlnAlaAsp----- 26
DQ 66380 TTTGGGCACCTCCAGTGACCGGGCTCGGGTCTGCACACCTCCCATGAGGAGATGTC 66439
QY 27 -----HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
DQ 66440 TTCTCACAATGAATCAAAAGCCACATCTGTGGTTCTATCTCTGGGACCAGGTCTGCAGA 66499
QY 40 ArgThrLeuAlaValProProTropIleGluTyrGlnHis 52
DQ 66500 GCTGGTTTAGTTCATCTCTTGGTGTACTCCACAGAAAT 66538

RESULT 13
AAI18766
ID AAI18766 standard; DNA; 591 BP.
XX AAI18766;
XX 12-OCT-2001 (first entry)
XX Probe #8699 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX
```

```
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 25; SEQ ID NO 8699; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 591 BP; 139 A; 145 C; 156 G; 151 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.0217 Length: 591
Score: 90.00 Matches: 21
Percent Similarity: 41.18% Conservative: 7
Best Local Similarity: 30.88% Mismatches: 18
Query Match: 25.50% Indels: 22
DB: 4 Gaps: 2

US-09-774-954-9 (1-61) x AAI18766 (1-591)
QY 2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
DQ 387 CCGAGGGCAGTTGGAGCCCGCGGGCTCTCAGACTACAGGGTGTGCGCCAGCGCCA 446
QY 21 PheGlyAsnGlnAlaAsp----- 26
DQ 447 TTTGGGCACCTCCAGTGACCGGGCTCGGGTCTGCACACCTCCCATGAGGAGATGTC 506
QY 27 -----HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
DQ 507 TTCTCACAATGAATCAAAAGCCACATCTGTGGTTCTATCTCTGGGACCAGGTCTGCAGA 566
QY 40 ArgThrLeuAlaValProProTrop 47
DQ 567 GCTGGTTTAGTTCATCTCTTGGTGTACTCCACAGAAAT 66538

RESULT 14
AAI28271
ID AAI28271 standard; DNA; 591 BP.
XX AAI28271;
XX 12-OCT-2001 (first entry)
XX Probe #18204 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX
```


THIS PAGE BLANK (ISPTO)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	123	34.8	474	2	T15511	hypothetical prote
2	78	22.1	354	2	I51266	opsin, green-sensi
3	67	19.0	355	2	A42347	opsin, green-sensi
4	66	18.7	355	2	I51319	RH2 opsin - green
5	64.5	18.3	219	2	S72721	amidotransferase h
6	64.5	18.3	223	2	B86968	conserved hypothet
7	64	18.1	212	2	D82431	ChbY family protei
8	62.5	17.7	603	2	A87342	conserved hypothet
9	62	17.6	349	2	B45229	opsin, green-sensi
10	62	17.6	353	1	JN0120	rhodopsin - Japane
11	61.5	17.4	349	2	A45229	opsin, green-sensi
12	60.5	17.1	767	2	D95101	SpoE family protei
13	60.5	17.1	767	2	E97969	cell division prot
14	59.5	16.9	198	2	C70570	hypothetical prote
15	59.5	16.9	432	2	T41004	hypothetical prote
16	59.5	16.9	569	2	E86357	probable cholesterol
17	59.5	16.9	585	2	S72824	cholesterol oxidas
18	59	16.7	2482	2	I48932	cation-independent
19	59	16.7	2483	1	I49617	insulin-like growt
20	58.5	16.6	430	2	AB2316	hypothetical prote
21	58.5	16.6	504	2	B40829	activin receptor i
22	58.5	16.6	528	2	C40829	activin receptor i
23	58	16.4	225	2	B88712	protein C17H12.10
24	58	16.4	365	2	A36047	Galactoside 2-alph
25	58	16.4	386	2	D69040	conserved hypothet
26	58	16.4	478	2	G71443	hypothetical prote
27	58	16.4	912	2	F71433	probable growth re
28	57.5	16.3	461	2	T09318	hypothetical prote
29	57.5	16.3	491	2	S76943	hypothetical prote

Search completed: October 25, 2005, 15:36:44
Job time : 8.64066 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 35.8234 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	61	11 US-09-774-954-9	Sequence 9, Appli
2	353	100.0	365	11 US-09-774-954-2	Sequence 2, Appli
3	353	100.0	388	14 US-10-301-822-169	Sequence 169, App
4	353	100.0	397	11 US-09-774-954-6	Sequence 6, Appli
5	295	83.6	61	11 US-09-774-954-3	Sequence 3, Appli
6	223	63.2	402	11 US-11-097-143-18162	Sequence 18162, A
7	218	61.8	343	11 US-09-774-954-17	Sequence 17, Appl
8	123	34.8	474	11 US-09-774-954-8	Sequence 8, Appli
9	120	34.0	150	18 US-10-820-474A-102	Sequence 102, App
10	101	28.6	28	11 US-09-774-954-14	Sequence 14, Appl
11	86	24.4	110	16 US-10-425-115-327031	Sequence 327031,

12	80.5	22.8	152	15	US-10-424-599-266118	Sequence 266118,
13	77.5	22.0	189	16	US-10-425-115-201329	Sequence 201329,
14	77.5	22.0	538	16	US-10-437-963-113919	Sequence 113919,
15	73.5	20.8	601	14	US-10-156-761-12551	Sequence 12551, A
16	64.5	18.3	223	15	US-10-282-122A-63705	Sequence 63705, A
17	63.5	18.0	191	15	US-10-282-122A-62907	Sequence 62907, A
18	63.5	18.0	580	15	US-10-369-493-8280	Sequence 8280, Ap
19	63	17.8	156	16	US-10-425-115-282566	Sequence 282566,
20	62.5	17.7	1271	16	US-10-437-963-180066	Sequence 180066,
21	62	17.6	412	16	US-10-856-499-2258	Sequence 2258, Ap
22	61.5	17.4	119	16	US-10-437-963-148755	Sequence 148755,
23	61.5	17.4	543	15	US-10-424-599-145078	Sequence 145078,
24	61.5	17.4	1262	16	US-10-437-963-117000	Sequence 117000,
25	61	17.3	75	16	US-10-437-963-198684	Sequence 198684,
26	61	17.3	117	16	US-10-425-115-235644	Sequence 235644,
27	61	17.3	146	15	US-10-424-599-233161	Sequence 233161,
28	60.5	17.1	741	17	US-10-472-928-1698	Sequence 1698, Ap
29	60.5	17.1	763	18	US-10-617-320-3610	Sequence 3610, Ap
30	60.5	17.1	767	9	US-09-815-242-13343	Sequence 13343, A
31	60.5	17.1	767	9	US-09-815-242-13598	Sequence 13598, A
32	60.5	17.1	767	10	US-09-769-787-96	Sequence 96, Appl
33	60.5	17.1	767	15	US-10-282-122A-73911	Sequence 73911, A
34	60.5	17.1	783	9	US-09-775-978-2	Sequence 2, Appli
35	60	17.0	545	15	US-10-424-599-174345	Sequence 174345,
36	60	17.0	557	15	US-10-425-114-49507	Sequence 49507, A
37	59.5	16.9	169	15	US-10-424-599-195444	Sequence 195444,
38	59.5	16.9	198	15	US-10-282-122A-61790	Sequence 61790, A
39	59.5	16.9	198	15	US-10-282-122A-62362	Sequence 62362, A
40	59.5	16.9	198	15	US-10-282-122A-64747	Sequence 64747, A
41	59.5	16.9	398	9	US-09-796-338A-17	Sequence 17, Appl
42	59.5	16.9	398	14	US-10-282-837-17	Sequence 17, Appl
43	59.5	16.9	398	14	US-10-225-567A-477	Sequence 477, App
44	59.5	16.9	398	14	US-10-145-586-17	Sequence 17, Appl
45	59.5	16.9	398	14	US-10-241-220-104	Sequence 104, App

ALIGNMENTS

RESULT 1

US-09-774-954-9
; Sequence 9, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-774-954-9
Query Match 100.0%; Score 353; DB 11; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.2e-37;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 60
DB 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 60
QY 61 H 61
DB 61 H 61
RESULT 2
US-09-774-954-2
; Sequence 2, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-774-954-2
Query Match 100.0%; Score 353; DB 11; Length 365;
Best Local Similarity 100.0%; Pred. No. 1.6e-36;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 60

```

```

DB 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 60
QY 61 H 61
DB 61 H 61
RESULT 3
US-10-301-822-169
; Sequence 169, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlagesel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029E2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 169
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-169
Query Match 100.0%; Score 353; DB 14; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.7e-36;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 60
DB 24 MPAGSWDPAGYLLYPCMGFRGNQADHFLGSLAFAKLNRTLVAVPPWIEYQHKKPPTNL 83
QY 61 H 61
DB 84 H 84
RESULT 4
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)

```


STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-Nov-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-Jan-1997
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-774-954-17
Query Match 61.8%; Score 218; DB 11; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.9e-19;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 NQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPFTNLH 61
DB 1 NQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPFTNLH 39
RESULT 8
US-09-774-954-8
Sequence 8, Application US/09774954
Publication No. US20040241845A1
GENERAL INFORMATION:
APPLICANT: Yang Wang, Michael W. Spellman
TITLE OF INVENTION: O-Fucosyltransferase
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/774,954
FILING DATE: 30-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-Nov-1997
APPLICATION NUMBER: 08/792,498
FILING DATE: 31-Jan-1997
ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-774-954-8
Query Match 34.8%; Score 123; DB 11; Length 474;
Best Local Similarity 67.6%; Pred. No. 6.1e-07;
Matches 23; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
QY 19 GRFGNQADHFLGSLAFKLNRTLAVPPWIEYQH 52
DB 166 GRFGNQVDQFLGVLAFKALDRTLVLPNFIEFKH 199
RESULT 9
US-10-820-474A-102
Sequence 102, Application US/10820474A
Publication No. US20050155089A1
GENERAL INFORMATION:
APPLICANT: LAL, PREETI
APPLICANT: TANG, Y. TOM
APPLICANT: GORGONE, GINA A.
APPLICANT: CORLEY, NEIL C.
APPLICANT: GUEGLER, KARL J.
APPLICANT: BAUGHN, MARIAH R.
APPLICANT: AKERBLOM, INGRID E.
APPLICANT: AU-YOUNG, JANICE
APPLICANT: YUE, HENRY
APPLICANT: PATTERSON, CHANDRA
APPLICANT: REDDY, ROOPA
APPLICANT: HILLMAN, JENNIFER L.
APPLICANT: BANDMAN, OLGA
TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
FILE REFERENCE: 039486-1568
CURRENT APPLICATION NUMBER: US/10/820,474A
CURRENT FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: 09/720,533
PRIOR FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: PCT/US99/14484
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: 60/090,762
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/094,983
PRIOR FILING DATE: 1998-07-31
PRIOR APPLICATION NUMBER: 60/102,686
PRIOR FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 269
SOFTWARE: PatentIn version 3.3
SEQ ID NO 102
LENGTH: 150
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY, misc feature
OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-102
Query Match 34.0%; Score 120; DB 18; Length 150;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYCFCMGR 20
DB 24 MPAGSWDPAGYLLYCFCMGR 43

```

RESULT 10
US-09-774-954-14
; Sequence 14, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-774-954-14
;
; Query Match 28.6% Score 101; DB 11; Length 28;
; Best Local Similarity 90.08; Pred. No. 1.8e-05;
; Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
; QY 1 MPAGSWDPAGYLLYPCWGR 20
; Db |||||
; 9-MPAGSWDPAGYLLYXPWGR 28
;
; RESULT 11
US-10-425-115-327031
; Sequence 327031, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 327031
; LENGTH: 110
; TYPE: DPT

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_115198C.1.pep
US-10-425-115-201329

Query Match      22.0%; Score 77.5; DB 16; Length 189;
Best Local Similarity 45.5%; Pred. No. 0.15;
Matches 20; Conservative 3; Mismatches 18; Indels 3; Gaps 2;

QY      10 GYLILCFMCRFGNQADHFLGSLAFKILNETLAVPPWIEYQHH 53
          ||||| : ||||| ||||| : |||||
DB      78 GGLWYAPHSG-FSNQVGELRNAVAARLLNRTLVVPPVLD--HH 118

RESULT 14
US-10-437-963-113919
; Sequence 113919, Application US/10437963
; Publication No. US20040123343A1
; OTHER INFORMATION:

```

; Sequence 113919, Application US/10437963
 ; Publication No. US20040123343A1
 GENERAL INFORMATION.

```

APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5322)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 113919
LENGTH: 538
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17661C.1.pep
US-10-437-963-113919

Query Match      22.0%; Score 77.5; DB 16; Length 538;
Best Local Similarity 44.2%; Pred. No. 0.47;
Matches 19; Conservative 4; Mismatches 17; Indels 3; Gaps 2;

QY          11 YLLYCPCMGKRGNCADHFLGSLAFAKLNRITLVAPPWIEYQH 53
           :                               :
bh         110 FIWTAPHSG-FSNCGVGLRNVAAVAALLNRTLVPVPVLDD--HH 149

```

```

RESULT 15
US-10-156-761-12551
: Sequence 12551, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 12551
: LENGTH: 601
: TYPE: PRT

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 10.0205 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGLYLPCMGCR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	353	100.0	61	3	US-08-978-741-9
2	353	100.0	61	3	US-09-333-729A-4
3	353	100.0	365	3	US-08-978-741-2
4	353	100.0	365	3	US-09-333-729A-3
5	353	100.0	397	3	US-08-978-741-6
6	353	100.0	397	3	US-09-333-729A-7
7	295	83.6	61	3	US-08-978-741-3
8	295	83.6	61	3	US-09-333-729A-5
9	223	63.2	399	4	US-09-270-767-45921
10	218	61.8	343	3	US-08-978-741-17
11	218	61.8	343	3	US-09-333-729A-13
12	123	34.8	474	3	US-08-978-741-8
13	123	34.8	474	3	US-09-333-729A-12
14	101	28.6	28	3	US-08-978-741-14
15	101	28.6	28	3	US-09-333-729A-14
16	63.5	18.0	205	4	US-09-540-236-3055
17	62.5	17.7	340	3	US-08-578-592-5
18	62.5	17.7	340	3	US-09-185-111-5
19	62	17.6	412	4	US-09-640-211A-2258
20	60.5	17.1	737	4	US-09-583-110-5075
21	60.5	17.1	763	4	US-09-107-433-3610
22	60.5	17.1	783	2	US-08-922-837-2
23	60.5	17.1	783	3	US-09-351-550-2
24	59.5	16.9	398	4	US-09-830-428A-5
25	59	16.7	786	4	US-09-543-681A-6379
26	58.5	16.6	315	4	US-09-902-540-16137
27	58	16.4	333	5	PCT-US91-00899-11

28	58	16.4	365	1	US-07-914-281-6	Sequence 6, Appli
29	58	16.4	365	1	US-08-393-246-6	Sequence 6, Appli
30	58	16.4	365	1	US-08-273-411-1	Sequence 1, Appli
31	58	16.4	365	1	US-08-525-058A-6	Sequence 6, Appli
32	58	16.4	365	1	US-08-395-800A-6	Sequence 6, Appli
33	58	16.4	365	1	US-08-395-800A-10	Sequence 10, Appli
34	58	16.4	365	2	US-08-696-731-6	Sequence 6, Appli
35	58	16.4	365	3	US-09-042-531-6	Sequence 2, Appli
36	58	16.4	365	3	US-09-151-592-2	Sequence 10, Appli
37	58	16.4	365	3	US-09-254-077A-10	Sequence 11, Appli
38	58	16.4	365	3	US-09-254-077A-11	Sequence 13, Appli
39	58	16.4	365	4	US-09-443-766-13	Sequence 12, Appli
40	58	16.4	365	5	PCT-US91-00899-12	Sequence 11353, A
41	58	16.4	399	4	US-09-949-016-11353	Sequence 64, Appli
42	57.5	16.3	461	3	US-09-457-046B-64	Sequence 25627, A
43	57.5	16.3	461	4	US-09-866-570B-64	Sequence 9626, Ap
44	56.5	16.0	288	4	US-09-252-991A-25627	
45	56.5	16.0	350	4	US-09-489-039A-9626	

ALIGNMENTS

RESULT 1
US-08-978-741-9
; Sequence 9, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear

US-08-978-741-9
Query Match 100.0%; Score 353; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.3e-40;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPAGSWDPAGLYLPCMGCRFGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNLH	60
Db	1	MPAGSWDPAGLYLPCMGCRFGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNLH	60
Qy			61 H 61

```

;
; LENGTH: 365 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-2

Query Match      100.0%; Score 353; DB 3; Length 365;
Best Local Similarity 100.0%; Pred. No. 2e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGREFGQADHFLGSLAFAKLNRTLVAVPPWIEYQHHKPPFTNL 60
Db 1 MPAGSWDPAGYLLYPCMGREFGQADHFLGSLAFAKLNRTLVAVPPWIEYQHHKPPFTNL 60
QY 61 H 61
Db 61 H 61

RESULT 4
US-09-333-729A-3
; Sequence 3, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 3
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-333-729A-3

Query Match      100.0%; Score 353; DB 3; Length 365;
Best Local Similarity 100.0%; Pred. No. 2e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCMGREFGQADHFLGSLAFAKLNRTLVAVPPWIEYQHHKPPFTNL 60
Db 1 MPAGSWDPAGYLLYPCMGREFGQADHFLGSLAFAKLNRTLVAVPPWIEYQHHKPPFTNL 60
QY 61 H 61
Db 61 H 61

RESULT 5
US-08-978-741-6
; Sequence 5, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svaboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: 08/792498
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-3

Query Match 100.0%; Score 353; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.2e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
DB 33 MPAGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 92

QY 61 H 61
DB 93 H 93

RESULT 6
US-09-333-729A-7
; Sequence 7, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid insert encoded protein.
; US-09-333-729A-7

Query Match 100.0%; Score 353; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 2.2e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 60
DB 33 MPAGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNL 92

QY 61 H 61
DB 93 H 93

RESULT 7
US-08-978-741-3
; Sequence 3, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
```

```

; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-978-741-3

Query Match 83.6%; Score 295; DB 3; Length 61;
Best Local Similarity 91.5%; Pred. No. 1.6e-32;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNLH 61
DB 3 AGSWDLAGYLLYXPXMGFRGNQADHFLGSLAFKLVRTLAVPPWIEYQHKKPPTNLH 61

RESULT 8
US-09-333-729A-5
; Sequence 5, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; APPLICANT: Spellman, Michael W.
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041P1-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Cricetulus Grieseus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 15, 17, 38
; OTHER INFORMATION: unknown amino acid
; US-09-333-729A-5

Query Match 83.6%; Score 295; DB 3; Length 61;
Best Local Similarity 91.5%; Pred. No. 1.6e-32;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDPAGYLLYCPGMRFGNQADHFLGSLAFKLNRTLAVPPWIEYQHKKPPTNLH 61
```

DB 3 AGSWDLAGLYLXPXMGFRGNQADHFLGSLAFAKLVETLAVPPWIEYQHKKPPTNLH 61

RESULT 9
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 45921
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-45921

Query Match 63.2%; Score 223; DB 4; Length 399;
Best Local Similarity 84.4%; Pred. No. 8.6e-22;
Matches 38; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 DPAGYLLYPCMGFRGNQADHFLGSLAFAKLVETLAVPPWIEYQ 51
DB 25 DPNGYLLYPCMGFRGNQADHFLGSLAFAKLVETLAVPPWIEYR 69

RESULT 10
US-08-978-741-17
; Sequence 17, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041p1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-17
Query Match 61.8%; Score 218; DB 3; Length 343;

Best Local Similarity 100.0%; Pred. No. 3.4e-21;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 NOADHFLGSLAFAKLVETLAVPPWIEYQHKKPPTNLH 61
DB 1 NOADHFLGSLAFAKLVETLAVPPWIEYQHKKPPTNLH 39

RESULT 11
US-09-333-729A-13
; Sequence 13, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
; APPLICANT: Wang, Yang
; TITLE OF INVENTION: O-Fucosyltransferase
; FILE REFERENCE: P1041PDI-Substitute
; CURRENT APPLICATION NUMBER: US/09/333,729A
; CURRENT FILING DATE: 1999-06-15
; PRIOR APPLICATION NUMBER: US 08/798,741
; PRIOR FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 13
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-333-729A-13

Query Match 61.8%; Score 218; DB 3; Length 343;
Best Local Similarity 100.0%; Pred. No. 3.4e-21;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 NOADHFLGSLAFAKLVETLAVPPWIEYQHKKPPTNLH 61
DB 1 NOADHFLGSLAFAKLVETLAVPPWIEYQHKKPPTNLH 39

RESULT 12
US-08-978-741-8
; Sequence 8, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

LENGTH: 474 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-978-741-8

Query Match 34.8%; Score 123; DB 3; Length 474;
Best Local Similarity 67.6%; Pred. No. 3.6e-08;
Matches 23; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 19 GRFGNQADHFLGSLAFKLNRTLVLPVPPWIEYQH 52
Db 166 GRFGNQVDQLGLVLAFAKALDRTLVLNPFIEFKH 199

RESULT 13
US-09-333-729A-12

Sequence 12, Application US/093333729A
Patent No. 6270987
GENERAL INFORMATION:
APPLICANT: Wang, Yang
APPLICANT: Spellman, Michael W.
TITLE OF INVENTION: O-Fucosyltransferase
FILE REFERENCE: P1041PDI-Substitute
CURRENT APPLICATION NUMBER: US/09/333, 729A
CURRENT FILING DATE: 1999-06-15
PRIOR APPLICATION NUMBER: US 08/798,741
PRIOR FILING DATE: 1997-11-26
NUMBER OF SEQ ID NOS: 21
SEQ ID NO 12
LENGTH: 474
TYPE: PRT
ORGANISM: Caenorhabditis Elegans
US-09-333-729A-12

Query Match 34.8%; Score 123; DB 3; Length 474;
Best Local Similarity 67.6%; Pred. No. 3.6e-08;
Matches 23; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 19 GRFGNQADHFLGSLAFKLNRTLVLPVPPWIEYQH 52
Db 166 GRFGNQVDQLGLVLAFAKALDRTLVLNPFIEFKH 199

RESULT 14
US-08-978-741-14

Sequence 14, Application US/08978741
Patent No. 6100076
GENERAL INFORMATION:
APPLICANT: Yang Wang, Michael W. Spellman
TITLE OF INVENTION: O-Fucosyltransferase
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,741
FILING DATE: 26-No. 6100076-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/792498
FILING DATE: 31
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044

REFERENCE/DOCKET NUMBER: P1041P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-978-741-14

Query Match 28.6%; Score 101; DB 3; Length 28;
Best Local Similarity 90.0%; Pred. No. 1.1e-06;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCWGR 20
Db 9 MPAGSWDPAGYLLYPCWGR 28

RESULT 15
US-09-333-729A-14

Sequence 14, Application US/093333729A
Patent No. 6270987
GENERAL INFORMATION:
APPLICANT: Wang, Yang
APPLICANT: Spellman, Michael W.
TITLE OF INVENTION: O-Fucosyltransferase
FILE REFERENCE: P1041PDI-Substitute
CURRENT APPLICATION NUMBER: US/09/333, 729A
CURRENT FILING DATE: 1999-06-15
PRIOR APPLICATION NUMBER: US 08/798,741
PRIOR FILING DATE: 1997-11-26
NUMBER OF SEQ ID NOS: 21
SEQ ID NO 14
LENGTH: 28
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Plasmid insert encoded protein.
NAME/KEY: unsure
LOCATION: 23, 25
OTHER INFORMATION: unknown amino acid
US-09-333-729A-14

Query Match 28.6%; Score 101; DB 3; Length 28;
Best Local Similarity 90.0%; Pred. No. 1.1e-06;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCWGR 20
Db 9 MPAGSWDPAGYLLYPCWGR 28

Search completed: October 25, 2005, 15:42:49
Job time : 11.0205 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 38.8296 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYPCMGCR.....LAVPPWIEYQHKKPPTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	353	100.0	61	2	Aaw80574 N-termina
2	353	100.0	365	2	Aaw80571 Human hea
3	353	100.0	388	8	Abm82223 Tumour-as
4	353	100.0	397	2	Aaw80573 Human O-f
5	295	83.6	61	2	Aaw80572 N-termina
6	241.5	68.4	417	4	Aam78377 Human pro
7	229.5	65.0	417	4	Aam79361 Human pro
8	223	63.2	402	4	Abb63790 Drosophil
9	218	61.8	343	3	Aaw80577 Partial h
10	120	34.0	150	3	Aay87325 Human sig
11	101	28.6	28	2	Aaw80575 N-termina
12	69.5	19.7	1561	6	Abb97437 Novel hum
13	64.5	18.3	223	6	Abu35781 Protein e
14	64	18.1	366	6	Abm69163 Phototrab
15	63.5	18.0	191	6	Abu34983 Protein e
16	63.5	18.0	205	8	Adl05369 M. catarr
17	63.5	18.0	580	8	Adn25627 Bacterial
18	62.5	17.7	340	2	Aar63066 Acetyl-es
19	62	17.6	412	3	Abb33303 Pinus rad
20	60.5	17.1	737	8	Adk48560 Streptoco
21	60.5	17.1	741	6	Abu01274 S. pneumo
22	60.5	17.1	763	8	Adr94975 Novel S.
23	60.5	17.1	767	4	Aau37750 Streptoco
24	60.5	17.1	767	4	Aau38005 Streptoco
25	60.5	17.1	767	6	Abu45987 Protein e

ALIGNMENTS

RESULT 1

AAW80574
ID AAW80574 standard; peptide; 61 AA.

XX AC AAW80574;

XX DT 16-DEC-1998 (first entry)

XX DE N-terminal amino acid sequence of human heart O-fucosyltransferase.

XX KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;

XX KM O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.

XX OS Homo sapiens.

XX PN WO9833924-A1.

XX PD 06-AUG-1998.

XX PF 17-DEC-1997; 97WO-US023401.

XX PR 31-JAN-1997; 97US-00792498.

XX PR 26-NOV-1997; 97US-00978741.

XX (GETH) GENENTECH INC.

XX Wang Y, Spellman MW;

XX WPI; 1998-437477/37.

XX Human O-fucosyltransferase able to glycosylate epidermal growth factor

XX Domains - useful for diagnosis and treatment of diseases involving

XX overexpression of the enzyme.

XX Claim 4; Page 69; 90pp; English.

XX This represents a the N-terminal sequence of the human heart O-fucosyltransferase that can glycosylate an epidermal growth factor (EGF) domain of a polypeptide with an activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity for the EGF domains, are used in diagnosis and treatment of conditions associated with overexpression of O-fucosyltransferase, to promote survival of sensory (retinal) neurons. Probes based on EGF domain polypeptide are used to detect gene amplification and expression. The expression can also be determined at the protein level using antibodies specific for O-fucosyltransferase

XX Sequence 61 AA;

SQ

Query Match 100.0%; Score 353; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 4.8e-40;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPFTNL 60
DB 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPFTNL 60

QY 61 H 61
DB 61 H 61

RESULT 2
ID AAW80571 standard; protein; 365 AA.
AC AAW80571;
DT 16-DEC-1998 (first entry)
DE Human heart O-fucosyltransferase.
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..61
FT /note= "N-terminal sequence claimed for in claim 4"
XX WO9833924-A1.
XX 06-AUG-1998.
XX 17-DEC-1997; 97WO-US023401.
XX 31-JAN-1997; 97US-00792498.
XX 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
XX Wang Y, Spellman MW;
XX WPI; 1998-437477/37.
XX N-PSDB; AAV65632.
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
XX domains - useful for diagnosis and treatment of diseases involving
XX overexpression of the enzyme.
XX Example 1; Fig 12A; 90pp; English.
XX This represents a human heart O-fucosyltransferase that can glycosylate
XX an epidermal growth factor (EGF) domain of a polypeptide with an
XX activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
XX mutants with increased affinity for the EGF domains, are used in
XX diagnosis and treatment of conditions associated with overexpression of O
XX -fucosyltransferase, to promote survival of sensory (retinal) neurons.
XX Probes based on EGF domain polypeptide are used to detect gene
XX amplification and expression. The expression can also be determined at
XX the protein level using antibodies specific for O-fucosyltransferase
XX Sequence 365 AA;
SQ

Query Match 100.0%; Score 353; DB 2; Length 365;
Best Local Similarity 100.0%; Pred. No. 3.8e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPFTNL 60

DB 1 MPAGSWDPAGYLLYPCPCMGREGNQADHFLGSLAFKLLNRTLAVPPWIEYQHHKPPFTNL 60

QY 61 H 61
DB 61 H 61

RESULT 3
ID ABM82223 standard; protein; 388 AA.
AC ABM82223;
DT 18-NOV-2004 (first entry)
DE Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic.
XX Homo sapiens.
XX WO2004030615-A2.
XX 15-APR-2004.
XX 29-SEP-2003; 2003WO-US028547.
XX 02-OCT-2002; 2002US-0414971P.
XX (GETH) GENENTECH INC.
XX Wu TD, Zhang Z, Zhou Y;
XX WPI; 2004-347921/32.
XX N-PSDB; ACN40746.
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX Claim 12; SEQ ID NO 5726; 7273pp; English.
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acids and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT polypeptide of the invention
SQ Sequence 388 AA;

Query Match 100.0%; Score 353; DB 8; Length 388;

Best Local Similarity 100.0%; Pred. No. 4.1e-39; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNL 60
24 MPAGSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNL 83

Db 61 H 61
84 H 84

RESULT 4
AAW80573
ID AAW80573 standard; protein; 397 AA.
XX AC
XX AAW80573;
DT 16-DEC-1998 (first entry)
DE Human O-fucosyltransferase sequence expressed by a plasmid insertion.
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Peptide 27-32
FT /note= "polyhistidine tag"
FT
XX WO9833924-A1.
XX 06-AUG-1998.
XX 17-DEC-1997; 97WO-US023401.
XX 31-JAN-1997; 97US-00792498.
XX 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
XX Wang Y, Spellman MW;
XX WPI; 1998-437477/37.
XX N-PSDB; AAV65633.
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX Example; Fig 13B; 90pp; English.
XX This represents the amino acid sequence of the human heart O-fucosyltransferase expressed by a plasmid insertion. The human O-fucosyltransferase can glycosylate an epidermal growth factor (EGF) domain of a polypeptide with an activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity for the EGF domains, are used in diagnosis and treatment of conditions associated with overexpression of O-fucosyltransferase, to promote survival of sensory (retinal) neurons. Probes based on EGF domain polypeptide are used to detect gene amplification and expression. The expression can also be determined at the protein level using antibodies specific for O-fucosyltransferase

Query Match 100.0%; Score 353; DB 2; Length 397;
Best Local Similarity 100.0%; Pred. No. 4.2e-39;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPAGSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNL 60

Db 33 MPAGSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNL 92

QY 61 H 61
93 H 93

Db

RESULT 5
AAW80572
ID AAW80572 standard; peptide; 61 AA.
XX AC
XX AAW80572;
DT 16-DEC-1998 (first entry)
DE N-terminal amino acid sequence of CHO O-fucosyltransferase.
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Misc-difference 15 /note= "unknown"
FT
FT Misc-difference 17 /note= "unknown"
FT
FT Misc-difference 38 /note= "unknown"
FT
XX WO9833924-A1.
XX 06-AUG-1998.
XX 17-DEC-1997; 97WO-US023401.
XX 31-JAN-1997; 97US-00792498.
XX 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
XX Wang Y, Spellman MW;
XX WPI; 1998-437477/37.
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX Claim 5; Page 69; 90pp; English.
XX This represents the N-terminal sequence of CHO O-fucosyltransferase. The enzyme can glycosylate an epidermal growth factor (EGF) domain of a polypeptide with an activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity for the EGF domains, are used in diagnosis and treatment of conditions associated with overexpression of O-fucosyltransferase, to promote survival of sensory (retinal) neurons. Probes based on EGF domain polypeptide are used to detect gene amplification and expression. The expression can also be determined at the protein level using antibodies specific for O-fucosyltransferase

Query Match 83.6%; Score 295; DB 2; Length 61;
Best Local Similarity 91.5%; Pred. No. 3.8e-32;
Matches 54; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AGSWDPAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNLH 61
3 AGSWDLAGYLLYPCPCMGFRGNQADHFLGSLAFKLLNRTLAVPPWIEYQHKKPPTNLH 61

```
RESULT 6
AAM78377
ID AAM78377 standard; protein; 417 AA.
XX
AC AAM78377;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1039.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
(PHYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
WPI; 2001-476283/51.
DR N-PSDB; AAK51510.
XX
Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX
Claim 20; Page 3276-3277; 6221pp; English.
XX
The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
sequence listing were missing at the time of publication
XX
Sequence 417 AA;
SQ
Query Match 68.4%; Score 241.5; DB 4; Length 417;
Best Local Similarity 76.7%; Pred. No. 6.8e-24;
Matches 46; Conservative 3; Mismatches 6; Indels 5; Gaps 1;
QY 7 DPAGYLLYCPCH-----GFGNQADHFLGSLAFAKLNRTLAVPPWIEYQHKKPPTNLH 61
Db 54 DMTATWTHAPSVEEGCGQRFNGQADHFLGSLAFAKLNRTLAVPPWIEYQHKKPPTNLH 113
RESULT 7
ABB63790
ID ABB63790 standard; protein; 402 AA.
```

```
AAM79361
ID AAM79361 standard; protein; 417 AA.
XX
AC AAM79361;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3007.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
(HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
WPI; 2001-476283/51.
DR N-PSDB; AAK52494.
XX
Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX
Claim 20; Page 229-230; 6221pp; English.
XX
The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
(AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
sequence listing were missing at the time of publication
XX
Sequence 417 AA;
SQ
Query Match 65.0%; Score 229.5; DB 4; Length 417;
Best Local Similarity 73.3%; Pred. No. 2.9e-22;
Matches 44; Conservative 4; Mismatches 7; Indels 5; Gaps 1;
QY 7 DPAGYLLYCPCH-----GFGNQADHFLGSLAFAKLNRTLAVPPWIEYQHKKPPTNLH 61
Db 54 DMTATWTHAPSVEEGCGQRFNGQADHFLGSLAFAKLNRTLAVPPWIEYQHKKPPTNLH 113
RESULT 8
ABB63790
ID ABB63790 standard; protein; 402 AA.
```

XX ABB63790;
AC
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 18162.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEXE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07893.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 402 AA;
Query Match 63.2%; Score 223; DB 4; Length 402;
Best Local Similarity 84.4%; Pred. No. 2.2e-21;
Matches 38; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 7 DPAGYLLYCFMGRFGNQADHFLGSLAFKLNRLTLAVPPWIEYQ 51
Db 28 DPNGYLYTCFMCGRFGNQADHFLGSLAFKLNRLTLILPWEYR 72
RESULT 9
AAW80577
ID AAW80577 standard; protein; 343 AA.
XX
AC AAW80577;
XX
XX 16-DEC-1998 (first entry)
DT
XX Partial human sequence of unknown function from a myeloblast cell line.
XX
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
XX Homo sapiens.
XX
XX WO9833924-A1.
PN
XX

PD 06-AUG-1998.
XX
XX 17-DEC-1997; 97WO-US023401.
XX
XX 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
XX (GETH) GENENTECH INC.
XX
XX Wang Y, Spellman MW;
PI
XX WPI; 1998-437477/37.
DR N-PSDB; AAV65634.
XX
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
XX Example 1; Page 62-68; 90pp; English.
XX
XX This represents a published partial human sequence of unknown function
CC from a myeloblast cell line. The invention provides a human heart O-
CC fucosyltransferase enzyme that can glycosylate an epidermal growth factor
CC (EGF) domain of a polypeptide with an activated O-fucose residue.
CC Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
CC for the EFG domains, are used in diagnosis and treatment of conditions
CC associated with overexpression of O-fucosyltransferase, to promote
CC survival of sensory (retinal) neurons. Probes based on EGF domain
CC polypeptide are used to detect gene amplification and expression. The
CC expression can also be determined at the protein level using antibodies
CC specific for O-fucosyltransferase
XX
XX Sequence 343 AA;
SQ
Query Match 61.8%; Score 218; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 8.6e-21;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 NOADHFLGSLAFKLNRLTLAVPPWIEYQHKKPPFTNLH 61
Db 1 NQADHFLGSLAFKLNRLTLAVPPWIEYQHKKPPFTNLH 39
RESULT 10
AAW87325
ID AAY87325 standard; protein; 150 AA.
XX
XX AAY87325;
XX
XX 11-MAY-2000 (first entry)
DT
XX Human signal peptide containing protein HSPP-102 SEQ ID NO:102.
DE
XX Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; neurotropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's disease; ovulatory defect;
KW muscular dystrophy.
XX
XX Homo sapiens.
OS
XX WO200000610-A2.
PN
XX
XX 06-JAN-2000.
PD
XX
XX 25-JUN-1999; 99WO-US014484.
PF
XX
XX 26-JUN-1998; 98US-0090762P.
PR 31-JUN-1998; 98US-0094983P.
PR

PR 01-OCT-1998; 98US-0102686P.
FR 11-DEC-1998; 98US-0112129P.
PA (INCY-) INCYTE PHARM INC.
XX Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX WPI; 2000-160673/14.
DR N-PSDB; AAZ98210.
XX New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT disease.
XX Claim 1; Page 225-226; 327pp; English.
XX AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nontropic, hepatotropic,
CC neuroprotective, cardiovascular and antidiabetic activities, and can be
CC used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC or ribozyme therapeutics, for detecting related sequences or genetic
CC variations, and for chromosomal mapping. HSPP are also used to raise
CC specific antibodies (Ab) and to screen for agonists and antagonists
CC (potential therapeutic agents). Ab are used to diagnose or monitor HSPP
CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC competitive drug screens, and for purification of HSPP from natural
CC sources
XX
SQ Sequence 150 AA;
Query Match 34.0%; Score 120; DB 3; Length 150;
Best Local Similarity 95.0%; Pred. No. 7.3e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCMGK 20
DB 24 MPAGSWDPAGYLLYPCMGK 43
RESULT 11
ID AAW80575 standard; protein; 28 AA.
XX AAW80575;
XX AAW80575;
XX 16-DEC-1998 (first entry)
XX N-terminal sequence of expressed human O-fucosyltransferase.
XX O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 23
FT /note= "unknown"

FT Misc-difference 25 /note= "unknown"
XX WO9833924-A1.
XX 06-AUG-1998.
XX 17-DEC-1997; 97WO-US023401.
XX 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX (GETH) GENENTECH INC.
XX Wang Y, Spellman MW;
PI WPI; 1998-437477/37.
XX Human O-fucosyltransferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX Example 1; Page 39; 90pp; English.
XX This represents the N-terminal amino acid sequence of the human heart O-
CC fucosyltransferase expressed by a plasmid insertion. The human O-
CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EGF
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 28 AA;
Query Match 28.6%; Score 101; DB 2;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MPAGSWDPAGYLLYPCMGK 20
DB 9 MPAGSWDPAGYLLYPCMGK 28
RESULT 12
ID ABB97437 standard; protein; 1561 AA.
XX ABB97437;
XX ABB97437;
XX 27-JUN-2002 (first entry)
XX Novel human protein SEQ ID NO: 705.
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX Homo sapiens.
OS WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
PA

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX WPI; 2002-292408/33.
 DR N-PSDB; ABN32623.
 XX An isolated polynucleotide for treating diseases associated with its
 PT encoded polypeptide such as cancer and multiple sclerosis.
 XX Claim 20; SEQ ID NO 705; 509pp; English.
 XX The present invention provides the protein and coding sequences of 444
 CC novel human proteins. These were isolated from expressed sequences tags
 CC (ESTs). They can be used to stimulate cell growth, to regulate
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
 CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
 CC rheumatoid arthritis, and to treat nervous system disorders e.g.
 CC Parkinson's disease. The present sequence is a protein of the invention
 XX Sequence 1561 AA;
 SQ
 Query Match 19.7%; Score 69.5; DB 5; Length 1561;
 Best Local Similarity 29.1%; Pred. No. 8.4; Mismatches 7; Indels 9; Gaps 2;
 Matches 16; Conservative 7;
 OY 5 SMDPAGYLLYCPQMGFGNQADHFLGSLAFKLLNRTLAVPP---WIEYQHHKPP 56
 DB 1342 TWEQRYLLWNPCSHFYQGF-----TFCPLKNVGCCLIGPDNIWFNIQRYESP 1390
 RESULT 13
 ID ABU35781
 XX ABU35781 standard; protein; 223 AA.
 XX AC ABU35781;
 XX 19-JUN-2003 (first entry)
 XX Protein encoded by Prokaryotic essential gene #21308.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Mycobacterium leprae.
 XX WO20027183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US0009107.
 XX 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wall L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA39651.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 63705; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 223 AA;
 SQ
 Query Match 18.3%; Score 64.5; DB 6; Length 223;
 Best Local Similarity 43.8%; Pred. No. 4.2;
 Matches 14; Conservative 4; Mismatches 11; Indels 3; Gaps 1;
 OY 21 FGNQADHFLGSLAFKLLNRTLAV---PPWIE 49
 DB 141 FGRQVDSFEGDIGFAGLVDPVRAVFIAPWVE 172
 RESULT 14
 ID ABM69163
 XX ABM69163 standard; protein; 366 AA.
 XX AC ABM69163;
 XX 20-NOV-2003 (first entry)
 XX Photorhabdus luminescens protein sequence #2260.
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX whooping cough.
 XX Photorhabdus luminescens.
 XX WO200294867-A2.
 XX 28-NOV-2002.
 XX 07-FEB-2002; 2002WO-IB003040.
 XX 07-FEB-2001; 2001FR-00001659.
 XX (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

